Title: Genomic and archaeological evidence suggest a dual origin of domestic dogs

Authors: Laurent A. F. Frantz[1†*], Victoria E. Mullin[2†], Maud Pionnier-Capitan[3,4], Ophélie Lebrasseur[1], Morgane Ollivier[3], Angela Perri[5], Anna Linderholm[1,6], Valeria Mattiangeli[2], Matthew D. Teasdale[2], Evangelos A. Dimopoulos[1,7], Anne Tresset[4], Marilyne Duffraisse[3], Finbar McCormick[8], László Bartosiewicz[9], Erika Gál[10], Éva A. Nyerges[10], Mikhail V. Sablin[11], Stéphanie Bréhard[4], Marjan Mashkour[4], Adrian Bălăşescu[12], Benjamin Gillet[3], Sandrine Hughes[3], Olivier Chassaing[3], Christophe Hitte[13], Jean-Denis Vigne[4], Keith Dobney[14], Catherine Hänni[3], Daniel G. Bradley[2*] and Greger Larson[1*]

Affiliations:

[1] The Palaeogenomics & Bio-Archaeology Research Network, Research Laboratory for Archaeology and History of Art, The University of Oxford, Oxford, UK.

[2] Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Dublin 2, Ireland.

[3] CNRS/ENS de Lyon, IGFL, UMR 5242 and French National Platform of Paleogenetics, PALGENE, Ecole Normale Supérieure de Lyon, 46 allée d'Italie, 69364 Lyon Cedex 07, France / Université Grenoble Alpes, Laboratoire d'Ecologie Alpine (LECA), F-38000, Grenoble, France.

[4] CNRS/MNHN/SUs – UMR 7209, Archéozoologie et Archéobotanique, Sociétés, Pratiques et Environnements, Département Ecologie et Gestion de la Biodiversité, 55 rue Buffon, 75005 Paris, France.

[5] Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany.

[6] Department of Anthropology, Texas A&M University, College Station, TX 77843-4352, USA.

[7] School of Biology, Aristotle University of Thessaloniki, Thessaloniki, Greece.

[8] School of Geography, Archaeology and Palaeoecology, Queen's University Belfast, University Road, Belfast, Northern Ireland, UK.

[9] Osteoarchaeological Research Laboratory, University of Stockholm, Stockholm, Sweden.

[10] Archaeological Institute, Research Centre for the Humanities, Hungarian Academy of Sciences, Budapest, Hungary.

[11] Zoological Institute RAS, Universitetskaya nab. 1, 199034 Saint-Petersburg, Russia

[12] The National Museum of Romanian History, 12 Calea Victoriei, 030026 Bucharest, Romania.

[13] Institut de Génétique et Développement de Rennes, CNRS-UMR6290, Université de Rennes 1, Rennes, France.

[14] Department of Archaeology, School of Geosciences, University of Aberdeen,St. Mary's, Elphinstone Road, AB24 3UF, UK.

[*] Corresponding authors: Laurent A. F. Frantz – laurent.frantz@arch.ox.ac.uk; Greger Larson – greger.larson@arch.ox.ac.uk; Dan G. Bradley - dbradley@tcd.ie

[†] Contributed equally

Abstract: The geographic and temporal origins of dogs remain controversial. Here, we generated genetic sequences from 59 ancient dogs and a complete (28x) genome of a late Neolithic dog (~4,800 calBP) from Ireland. Our analyses revealed a deep split separating modern East Asian and Western Eurasian dogs. Surprisingly, the date of this divergence (~14,000-6,400 years ago) occurs commensurate or several millennia after the first appearance of dogs in Europe and East Asia. Additional analyses of ancient and modern mitochondrial DNA revealed a sharp discontinuity in haplotype frequencies in Europe. Combined, these results suggest that dogs may have been domesticated independently in Eastern and Western Eurasia from distinct wolf populations. East Eurasia dogs were then possibly transported alongside people where they partially replaced European Palaeolithic dogs.

One Sentence Summary: Genomics and archaeology reveal both a possible dual origin of domestic dogs and a subsequent translocation of East Asian dogs into Europe.

Main Text: Dogs were the first domestic animal and the only animal domesticated prior to the advent of settled agriculture (*1*). Despite their importance in human history, no consensus has emerged with regard to their geographic and temporal origins, or whether dogs were domesticated just once or independently on more than one occasion. Though several claims have been made for an initial appearance of dogs in the early Upper Palaeolithic (~30,000 years ago; *e.g. 2*), the first remains confidently assigned to dogs appear in Europe ~15,000 years ago and in Far East Asia over 12,500 years ago (*1*, *3*). While archaeologists remain open to the idea that there was more than one geographic origin for dogs *(e.g.* (*4*, *5*), most genetic studies have concluded that dogs were likely domesticated just once (*6*) – disagreeing on whether this occurred in Europe (*7*), Central Asia (*8*), or East Asia (*9*).

Recent palaeogenetic studies have had a tremendous impact on our understanding of early human evolution (*e.g.* (*10*, *11*)). Here we apply a similar approach to reconstruct the evolutionary history of dogs. We generated 59 ancient mtDNA sequences from European dogs (from 14,000 to 3,000 years ago) as well as a high coverage nuclear genome (~28x) of an ancient dog ~4,800 calBP (*12*) from the Neolithic passage grave complex of Newgrange (*Sí an Bhrú*) in Ireland. We combined our ancient sample with 80 modern publically available full genome sequences and 605 modern dogs (including village dogs and 48 breeds) genotyped on the 170k HD SNP array (*12*).

We first assessed characteristics of the Newgrange dog by typing SNPs associated with specific phenotypic traits and by inferring its level of inbreeding, compared to other breed and village dogs (*12*). Our results suggest that the degree of artificial selection and controlled breeding during the Neolithic was similar to that observed in modern free-living dogs. In addition, the Newgrange dog did not possess variants associated with modern breed-defining traits including hair length or coat color. And though this dog was likely able to digest starch less efficiently than modern dogs, it was more efficient than wolves (*12*).

A phylogenetic analysis, based on 170k SNPs revealed a deep split separating the modern Sarloos breed from other dogs (Fig. 1a). This breed - created in the 1930s in the Netherlands - involved breeding German Shepherds with captive wolves (*13*), thus explaining the breed's topological placement. Interestingly, the second deepest split (evident on the basis of both the 170K SNP panel – Fig 1a - and genome-wide SNPs - Fig. S4) separates modern East Asian and Western Eurasian (Europe and the Middle East) dogs. Moreover, the Newgrange dog clusters tightly with Western Eurasian dogs. We used Principal Component Analysis (PCA), D-statistics and *TreeMix* (*12*) to further test this pattern. Each of these analyses unequivocally placed the Newgrange dog with modern European dogs (Figs. S5, S6, S7). These findings demonstrate that the node separating the East Asian and Western Eurasian clades is older than the Newgrange individual; directly radiocarbon dated to ~4,800 years ago.

Other nodes leading to multiple dog populations and breeds (including the basal breeds (*1*) such as Greenland Sledge dogs or Siberian Husky; Fig. 1a) are poorly supported, suggesting that these breeds likely possess mixed ancestry from both Western Eurasian and East Asian dog lineages. To further assess the robustness of the deep split and those nodes associated with the potentially admixed lineages, we defined Western Eurasian and East Asian "core" groups (Fig. 1a) supported by the strength of the node leading to each cluster (*12*). We then used D-statistics to assess the affinity of each population to either Western Eurasian or East Asian core groups (*12*). The results of this analysis again revealed a clear East-West geographic pattern across Eurasia associated with the deep phylogenetic split (Fig. 1b). Breeds such as the Eurasier, Greenland Sledge dogs and Siberian Huskies (all basal breeds from Northern regions(*1*)), however, possess strong signatures of admixture with the East Asian core samples (Fig. S11), as do populations sampled in East Asia that clustered alongside Western Eurasian dogs (*e.g.* Papua New Guinean village dog; Fig. 1a).

We used the Multiple Sequentially Markovian Coalescent (*MSMC*)(*12*, *14*) to reconstruct the population history of East Asian and Western Eurasia dogs. An analysis of individual high coverage genomes demonstrated a long, shared population history between the Newgrange dog and modern dogs from both Western Eurasia and East Asia (Fig. S15). A reconstruction using two genomes per group improved the resolution for recent time periods (Fig. 2a) and revealed a bottleneck in the Western Eurasian population, following its divergence from the East Asian core. A similar bottleneck observed in non-African human populations has been interpreted as a signature of a migration out of Africa (*15*). We therefore speculate that the analogous bottleneck observed in our dataset could be the result of a divergence and subsequent migration from east to west; supporting suggestions drawn from recent analyses of modern dog genomes (*8*, *9*, *16*).

We then used *MSMC* to compute divergence times as a mean to assess the time frame of the shared population history among dogs, and between dogs and wolves. To obtain reliable time estimates, we used the radiocarbon age of the Newgrange dog to calibrate the mutation rate for dogs (*12*)(Fig. S16). This resulted in a mutation rate estimate of between 0.3×10^{-8} and 0.45×10^{-8} per generation - similar to that obtained with an ancient grey wolf genome (*17*). Using this mutation rate, we calculated the divergence time between the two modern Russian wolves (*18*) used in this study and the modern dogs to be 60,000-20,000 years ago (Fig. S17; Fig. 2b). Importantly, this date should not be interpreted as a time frame for domestication, since the wolves we examined may not have been closely related to the population that gave rise to dogs (*6*).

These analyses also suggested that the divergence between the East Asian and Western Eurasian core groups (~14,000-6,400 years ago) occurred commensurate, or several millennia after the earliest known appearance of domestic dogs in both Europe (>15,000 years) and East Asia (>12,500 years) (*1*) (Figs. S17, 2b). In addition, admixture signatures from wolves into Western Eurasian dogs most likely pushed this estimated time of divergence deeper into the past (*12*) meaning that the expected time of divergence between East and Western cores is likely younger than our estimate. These results imply that indigenous populations of dogs were already present in Europe and East Asia during the Palaeolithic (prior to this genomic divergence). Under this hypothesis, this early indigenous dog population in Europe was replaced (at least partially) by the arrival of East Eurasian dogs.

To investigate this potential replacement, we sequenced and analyzed 59 hyper-variable mtDNA fragments from ancient dogs spread across Europe and combined those with 167 modern sequences (*12*). Each sequence was then assigned to one of four major well-supported haplogroups (A-D) (*19*). While the majority of ancient European dogs belonged to either haplogroup C or D (63% and 20%, respectively), most modern European dogs possess sequences within haplogroups A and B (64 and 22% respectively) (Fig. 2c, d, e). Using simulations, we showed that this finding cannot be explained by drift alone (*12*). Instead, this pattern arose from clear turnover in the mitochondrial ancestry of European dogs, most likely as a result of an arrival of East Asian dogs. This migration led to a partial replacement of ancient dog lineages in Europe that were present by at least 15,000 years ago (*1*).

Though the mtDNA turnover is obvious, the nuclear signature reveals an apparent long-term continuity. Assessments of ancestry in humans have demonstrated that major (nuclear) turnovers can be difficult to detect without samples from the admixing population (*11*). A genome-wide PCA analysis revealed that PC2 clearly discriminates the Newgrange dog from other modern dogs (Fig. S8), suggesting that this individual possessed ancestry from an unsampled population.

Our MSMC analysis reveals that the population split between the Newgrange dog and the East Asian core (as measured by cross coalescence rate [CCR]) is older (on average) than the split between modern Western Eurasian and East Asian lineages (Fig. 2b). Simulations suggest that this pattern could be explained by a partial replacement model in which the Newgrange dog retained a degree of ancestry from an outgroup population (Fig. S20a,b), that was different from modern wolves (*12*). Alternatively, this pattern could also be explained by secondary gene flow from Asian dogs into modern European dogs (Fig. S20c). Nevertheless, simulations show that secondary gene flow has a smaller effect on CCR than the partial replacement model (Fig. S20b,d). Moreover, secondary gene flow cannot explain the placement of the Newgrange dog on our genome-wide PCA (Fig. S8). Overall, these observations are consistent with a scenario in which the Newgrange dog retained a degree of ancestry from an ancient canid population that falls outside of the variation of modern dogs, but that is also different from modern wolves. This pattern also suggests that the replacement of European indigenous Palaeolithic dogs may not have been complete.

To assess the consilience between our results and the archaeological record, we compiled evidence for the earliest dog remains across Eurasia (Fig. 3a). We found that while dogs are present at sites as old as 12,500 years in Eastern Eurasia (China, Kamchatka and East Siberia) and 15,000 years in Western Eurasia (Europe and Near East) dog remains older than 8,000 years have yet to be recovered in Central Eurasia (Fig. 3a; Table S7). Combined with our DNA

analyses, this observation suggests that two distinct populations of dogs were present in Eastern and Western Eurasia during the Palaeolithic.

The establishment of these populations is consistent with two scenarios: a single origin of Eurasian dogs followed by early transportation, founder effects, isolation and drift, or two independent domestication processes on either side of Eurasia. In the first scenario, the archaeological record should reveal a temporal cline of the first appearance of dogs across Eurasia stemming from a single source. Given the current lack of dog remains prior to 8,000 years ago in Central Eurasia, a scenario involving a single origin followed by an early transportation seems less likely.

Given our combined results, we suggest the following hypothesis: two genetically differentiated and potentially extinct wolf populations in Eastern (*8*, *9*) and Western Eurasia (*7*) may have been independently domesticated prior to the advent of settled agriculture (Fig. 3a). The eastern dog population then dispersed westward alongside humans, between 6,400 and 14,000 years ago, into Western Europe (*10*, *11*, *20*) whereupon they partially replaced an indigenous Palaeolithic dog population. Our hypothesis reconciles previous studies that have suggested domestic dogs originated in East Asia (*9*, *19*) and Europe (*7*). For numerous reasons, the null hypothesis should be that individual animal species were domesticated just once (*21*). The combined genetic and archaeological results presented here, however, suggest that dogs, like pigs (*22*), may have been domesticated twice. Nevertheless, given the complexity of the evolutionary history of dogs and uncertainties related to mutation rates, generation times and the incomplete nature of the archaeological record, our scenario remains hypothetical. Genome sequences derived from ancient Eurasian dogs and wolves will provide the necessary means to assess whether dog domestication occurred more than once.

References and Notes:

1. G. Larson *et al.*, *Proc. Natl. Acad. Sci.* **109**, 8878-83 (2012).
2. M. Germonpré, M. Lázničková-Galetová, M. V. Sablin, *J. Archaeol. Sci.* **39**, 184–202 (2012).
3. M. Pionnier-Capitan *et al.*, *J. Archaeol. Sci.* **38**, 2123–2140 (2011).
4. T. Dayan, *J. Archaeol. Sci.* **21**, 633–640 (1994).
5. M. Ollivier *et al.*, *PLoS One*. **8**, e75110 (2013).
6. A. H. Freedman *et al.*, *PLoS Genet.* **10**, e1004016 (2014).
7. O. Thalmann *et al.*, *Science*. **342**, 871–4 (2013).
8. L. M. Shannon *et al.*, *Proc. Natl. Acad. Sci. U. S. A.* **112**, 13639–13644 (2015).
9. G.-D. Wang *et al.*, *Cell Res.* **26**, 21-33 (2015).
10. M. E. Allentoft *et al.*, *Nature*. **522**, 167–172 (2015).
11. W. Haak *et al.*, *Nature*. **522**, 207–211 (2015).
12. Supplementary Material.
13. D. Morris, *Dogs: The Ultimate Dictionary of Over 1,000 Dog Breeds* (Trafalgar Square, 2008).

14. S. Schiffels, R. Durbin, *Nat. Genet.* **46**, 919–25 (2014).
15. H. Li, R. Durbin, *Nature*. **475**, 493–6 (2011).
16. M. Pilot *et al.*, *Proc. Biol. Sci.* **282**, 20152189– (2015).
17. P. Skoglund, E. Ersmark, E. Palkopoulou, L. Dalén, *Curr. Biol.* **25**, 1515–1519 (2015).
18. G. Wang *et al.*, *Nat. Commun.* **4**, 1860 (2013).
19. P. Savolainen, Y. Zhang, J. Luo, J. Lundeberg, T. Leitner, *Science*. **298**, 1610–3 (2002).
20. L. M. Cassidy *et al.*, *Proc. Natl. Acad. Sci.* **113**, 201518445 (2015).
21. G. Larson, J. Burger, *Trends Genet.* **29**, 197–205 (2013).
22. L. A. F. Frantz *et al.*, *Nat. Genet.* **47**, 1141–1148 (2015).
23. C. Gamba *et al.*, *Nat. Commun.* **5**, 5257 (2014).
24. L. H. van Wijngaarden-Bakker, *Proc. R. Irish Acad. Sect. C Archaeol. Celt. Stud. Hist. Linguist. Lit.* **86C**, 17–111 (1986).
25. M. J. O'Kelly, R. M. Cleary, D. Lehane, *Newgrange, Co. Meath, Ireland: The Late Neolithic/Beaker Period Settlement* (B.A.R., 1983).
26. F. McCormick, in *The Holocene History of the European Vertebrate Fauna: Modem Aspects of Research*, N. Benecke, Ed. (1998).
27. P. Reimer, *Radiocarbon*. **55**, 1869–1887 (2013).
28. S. Pääbo *et al.*, *Annu. Rev. Genet.* **38**, 645–679 (2004).
29. L. Orlando *et al.*, *Genome Res.* **21**, 1705–19 (2011).
30. J. T. Vilstrup *et al.*, *PLoS One*. **8**, e55950 (2013).
31. M. Meyer, M. Kircher, *Cold Spring Harb. Protoc.* **2010**, pdb.prot5448 (2010).
32. M. Martin, *EMBnet.journal*. **17**, 10 (2011).
33. H. Li, R. Durbin, *Bioinformatics*. **25**, 1754–60 (2009).
34. K. Lindblad-Toh *et al.*, *Nature*. **438**, 803–19 (2005).
35. M. Schubert *et al.*, *Nat. Protoc.* **9**, 1056–82 (2014).
36. H. Li *et al.*, *Bioinformatics.* **25**, 2078-9 (2009).
37. A. McKenna *et al.*, *Genome Res.* **20**, 1297–1303 (2010).
38. K. Okonechnikov, A. Conesa, F. García-Alcalde, *Bioinformatics*. **32**, 292-4 (2015).
39. H. Jónsson, A. Ginolhac, M. Schubert, P. L. F. Johnson, L. Orlando, *Bioinformatics*. **29**, 1682–4 (2013).
40. B. Bai *et al.*, *Nucleic Acids Res.* **43**, 777-83 (2014).
41. A. Auton *et al.*, *Science*. **336**, 193-8 (2012).
42. A. R. Quinlan, I. M. Hall, *Bioinformatics*. **26**, 841–2 (2010).
43. E. Han, J. S. Sinsheimer, J. Novembre, *Mol. Biol. Evol.* **31**, 723–35 (2014).

44. A. Vaysse *et al.*, *PLoS Genet.* **7**, e1002316 (2011).
45. M. L. Speir *et al.*, *Nucleic Acids Res.*, **44** 717–25 (2016).
46. S. Purcell *et al.*, *Am. J. Hum. Genet.* **81**, 559–75 (2007).
47. E. Paradis, J. Claude, K. Strimmer, *Bioinformatics*. **20**, 289–290 (2004).
48. N. Patterson, A. L. Price, D. Reich, *PLoS Genet.* **2**, e190 (2006).
49. J. K. Pickrell, J. K. Pritchard, *PLoS Genet.* **8**, e1002967 (2012).
50. E. Y. Durand, N. Patterson, D. Reich, M. Slatkin, *Mol. Biol. Evol.* **28**, 2239–2252 (2011).
51. N. Patterson *et al.*, *Genetics*. **192**, 1065–93 (2012).
52. S. R. Browning, B. L. Browning, *Am. J. Hum. Genet.* **81**, 1084–97 (2007).
53. Z. Fan *et al.*, *Genome Res.* **26**, 163-73 (2016).
54. H. Angleby, P. Savolainen, *Forensic Sci. Int.* **154**, 99–110 (2005).
55. A. Ardalan *et al.*, *Ecol. Evol.* **1**, 373–85 (2011).
56. A. R. Boyko *et al.*, *Proc. Natl. Acad. Sci. U. S. A.* **106**, 13903–8 (2009).
57. S. K. Brown *et al.*, *PLoS One*. **6**, e28496 (2011).
58. S. Castroviejo-Fisher, P. Skoglund, R. Valadez, C. Vilà, J. A. Leonard, *BMC Evol. Biol.* **11**, 73 (2011).
59. V. Muñoz-Fuentes, C. T. Darimont, P. C. Paquet, J. A. Leonard, *Conserv. Genet.* **11**, 547–556 (2009).
60. N. Okumura, N. Ishiguro, M. Nakano, A. Matsui, M. Sahara, *Anim. Genet.* **27**, 397–405 (2009).
61. M. C. R. Oskarsson *et al.*, *Proc. Biol. Sci.* **279**, 967–74 (2012).
62. J.-F. Pang *et al.*, *Mol. Biol. Evol.* **26**, 2849–64 (2009).
63. P. Savolainen, T. Leitner, A. N. Wilton, E. Matisoo-Smith, J. Lundeberg, *Proc. Natl. Acad. Sci. U. S. A.* **101**, 12387–90 (2004).
64. K. Tsuda, Y. Kikkawa, H. Yonekawa, Y. Tanabe, *Genes Genet. Syst.* **72**, 229–38 (1997).
65. J. A. Leonard *et al.*, *Science*. **298**, 1613–6 (2002).
66. M. F. Deguilloux, J. Moquel, M. H. Pemonge, G. Colombeau, *J. Archaeol. Sci.* **36**, 513–519 (2009).
67. F. Verginelli *et al.*, *Mol. Biol. Evol.* **22**, 2541–51 (2005).
68. M. Pionnier-Capitan, thesis, ENS Lyon (2010).
69. O. Lebrasseur, thesis, Durham University (2014).
70. N. Rohland, M. Hofreiter, *Nat. Protoc.* **2**, 1756–62 (2007).
71. A. Cooper, *Science (80-.).* **289**, 1139b–1139 (2000).
72. M. T. P. Gilbert, H.-J. Bandelt, M. Hofreiter, I. Barnes, *Trends Ecol. Evol.* **20**, 541–4

(2005).

73. K. S. Kim, S. E. Lee, H. W. Jeong, J. H. Ha, *Mol. Phylogenet. Evol.* **10**, 210–20 (1998).

74. M.-S. Peng *et al.*, *Mol. Ecol. Resour.* **15**, 1238–42 (2015).

75. R. C. Edgar, *Nucleic Acids Res.* **32**, 1792–7 (2004).

76. J. Leigh, PopART (Population Analysis with Reticulate Trees) (2015).

77. R. R. Hudson, *Bioinformatics*. **18**, 337–338 (2002).

78. G. K. Chen, P. Marjoram, J. D. Wall, *Genome Res.* **19**, 136–42 (2009).

79. D. H. Alexander, J. Novembre, K. Lange, *Genome Res.* **19**, 1655–64 (2009).

80. M. Boudadi-Maligne, J.-B. Mallye, M. Langlais, C. Barshay-Szmidt, *PALEO*, 39–54 (2012).

81. R. J. Losey *et al.*, *PLoS One*. **8**, e63740 (2013).

82. N. Benecke, *J. Archaeol. Sci.* **14**, 31–49 (1987).

83. M. Germonpré *et al.*, *J. Archaeol. Sci.* **36**, 473–490 (2009).

84. R. Flad, *J. East Asian Archaeol.* **3**, 23–51 (2001).

85. L. Barton *et al.*, *Proc. Natl. Acad. Sci. U. S. A.* **106**, 5523–8 (2009).

86. A. Lasota-Moskalewska, K. Szymczak, M. Khudzhanazarov, *Archaeol. Balt.* **11** (2009).

87. A. Razzokov. *Sarazm. Dushanbe: Institute of History, Archaeology, and Ethnography. Dushanbe, Tajikistan: Academy of Sciences of Tajikistan A. Donish History, Archaeology and Ethnographic Institute.* (2008).

88. H. Bocherens, M. Mashkour, D. Billiou, *Environ. Archaeol.* **33**, 253-64 (2013).

89. A. K. Kasparov, *Paléorient*. **22**, 161–167 (1996).

90. M. Frachetti, N. Benecke, *Antiquity*. **83**, 1023–1037 (2015).

91. C. Chang, N. Benecke, F. Grigoriev, A. Rosen, P. Tourtellotte, *Antiquity.* **77**, 298-312 (2003)

92. V. Bakhshaliev, in *The Archaeology of Nakhichevan. Ten Years of New Discoveries* (Ege Yayinlari, Istanbul, 1997).

93. R. Berthon *et al.*, *Environ. Archaeol.* **18**, 191-200 (2013).

94. S. J. Pawankar and P. K. Thomas, *Bulletin of the Deccan College Research Institute* **56**, 363-370 (1996).

95. B. & R. Allchin, F. R. Allchin, in *The Birth of Indian Civilization: India and Pakistan Before 500 B.C.* (1968).

96. R. V. Joshi, *Archaeol. Surv. India, New Delhi, India*, 12–16 (1961).

97. P. K. Thomas, in *The Walking larder: patterns of domestication, pastoralism, and predation* (Routledge, 2014).

98. B. P. Sahu, *From hunters to breeders: faunal background of early India* (Anamika

Prakashan, 1988).

99. N. Bhola, G. V. S. Rao, *Animal remains from Lothal excavations* (Zoological Survey of India, 1962).

100. R. B. . Sewell, B. S. Guha, in *Mohenjo-Daro and the Indus* (1931).

101. G. Stein, in *Early Animal Domestication and Its Cultural Context*, P. J. Crabtree, D. V. Campana, K. Ryan, Eds. (UPenn Museum of Archaeology, 1989).

102. Telegin D.Ya., *redno-Stogivska kultura epokhi midi.* (1973).

103. K. N. Wilkinson *et al.*, *J. F. Archaeol.* **37**, 20-33 (2013).

104. R. D. Barnett, W. Watson, *Iraq*. **14**, 132 (1952).

105. M. D. Frachetti, N. Benecke, A. N. Mar'yashev, P. N. Doumani, *World Archaeol.* **42**, 622–646 (2010).

106. P. M. Dolukhanov, in *Hunters in transition: Mesolithic societies of temperate Eurasia and their transition to farming* (1986).

107. G. Matyushin, *The mesolithic and neolithic in the southern Urals and central Asia.* (1986).

108. N. D. Ovodov *et al.*, *PLoS One*. **6**, e22821 (2011).

109. M. Germonpré, M. Lázničková-Galetová, R. J. Losey, J. Räikkönen, M. V. Sablin, *Quat. Int.* **359-360**, 261–279 (2015).

110. M.-A. Garcia, *Bull. la Société préhistorique française*. **102**, 103–108 (2005).

111. M. V. Sablin, G. A. Khlopachev, *Curr. Anthropol.* **43**, 795–799 (2002).

112. A. Perri, *J. Archaeol. Sci.* **68**, 1–4 (2016).

113. S. J. Crockford, Y. V. Kuzmin, *J. Archaeol. Sci.* **39**, 2797–2801 (2012).

114. M. Boudadi-Maligne, G. Escarguel, *J. Archaeol. Sci.* **45**, 80–89 (2014).

115. A. G. Drake, M. Coquerelle, G. Colombeau, *Sci. Rep.* **5**, 8299 (2015).

116. D. P. Howrigan, M. A. Simonson, M. C. Keller, *BMC Genomics*. **12**, 460 (2011).

117. E. Axelsson *et al.*, *Nature*. **495**, 360–4 (2013).

118. E. Cadieu *et al.*, *Science*. **326**, 150–3 (2009).

119. H. G. Parker *et al.*, *Science*. **325**, 995–8 (2009).

Acknowledgments: Raw reads of the Newgrange dog have been deposited at the European Nucleotide Archive (ENA) with project number: PRJEB13070. Mitochondrial sequences as well as genotype files (in plink format) were deposited on Dryad (http://datadryad.org/review?doi=doi:10.5061/dryad.8gp06). We would like to thank G. Wang, J. Schraiber, L. Orlando, L. Dalén, R.E. Green, P. Savolainen, E. Loftus for their valuable input. We are also grateful to A. Osztás and I. Zalai-Gaál R.-M. Arbogast, A. Beeching, A. Boroneant, O. Lecomte, S. Madeleine, C. & D. Mordant, M. Patou-Mathis, P. Pétrequin, L. Salanova, J. Schibler, A. Tsuneki, F. Valla for providing archaeological material. L.A.F.F., O.L., A.L. and

G.L. were supported by a European Research Council grant (ERC-2013-StG-337574-UNDEAD) and Natural Environmental Research Council grants (NE/K005243/1 and NE/K003259/1). L.A.F.F. was supported by a Junior Research Fellowship (Wolfson College, University of Oxford). V.E.M, V.M , M.D.T. and the sequencing of the Newgrange dog genome were funded by ERC Investigator grant 295729-CodeX awarded to D.G.B. We would also like to acknowledge the National Museum of Ireland for providing the petrous bone of the Newgrange dog and the Science Foundation Ireland Award 12/ERC/B2227 and Trinseq. AB was supported by a Romanian National Authority for Scientific Research (PN-II-ID-PCE-2011-3-1015). The work at ENS Lyon and at MNHN Paris was also supported by PURINA-NESTLE. The authors declare no conflict of interest.

Fig. 1: Deep split between East Asian and Western Eurasian dogs.*a.* A neighbour-joining tree (with bootstrap values) based on Identity by State (*12*) of 605 dogs. Red and yellow clades represent the East Asian and Western Asian core groups respectively (*12*). ***b.*** A map showing the location and relative proportion of ancestry (mean D-values) of dogs (Fig. S10). Positive values (red) indicate that the population shares more derived alleles with the East Asian core while negative values (yellow) indicate a closer association with the Western Eurasian core.

Fig. 2: Effective population size, divergence times and mtDNA. *a*. Effective population size through time of East and Western Eurasian dogs and wolves with MSMC. ***b*.** Cross-coalescence rate (CCR) per year for each population pair in Fig. 2a. The CCR represents the ratio of within and between population coalescence rates (CR). The ratio measures the age and pace of divergence between two populations. Values close to 1 indicate that both within and between CR are equal meaning the two populations have not yet diverged. Values close to 0 indicate that the populations have completely diverged. ***c*.** Bar plot representing the proportion of mtDNA haplogroups at different time periods. ***d.*** Locations of archaeological sites with haplogroup proportions. ***e.*** Location of modern samples with haplogroup proportions.

Fig. 3: Archaeological evidence for the first appearance of dogs across Eurasia and a model of dog domestication. *a.* Map representing the geographic origin and age of the oldest archaeological dog remains in Eurasia (*12*). ***b.*** A suggested model of dog domestication under the dual origin hypothesis. An initial wolf population split into East and West Eurasian wolves that were then domesticated independently before going extinct (as indicated by the † symbol). The Western Eurasian dog population (European) was then partially replaced by a human-mediated translocation of Asian dogs at least 6,400 years ago, a process that took place gradually after the arrival of the eastern dog population.

Supplementary Materials:

Materials and Methods

Figs. S1-S29

Tables S1-S7

References (23-110)

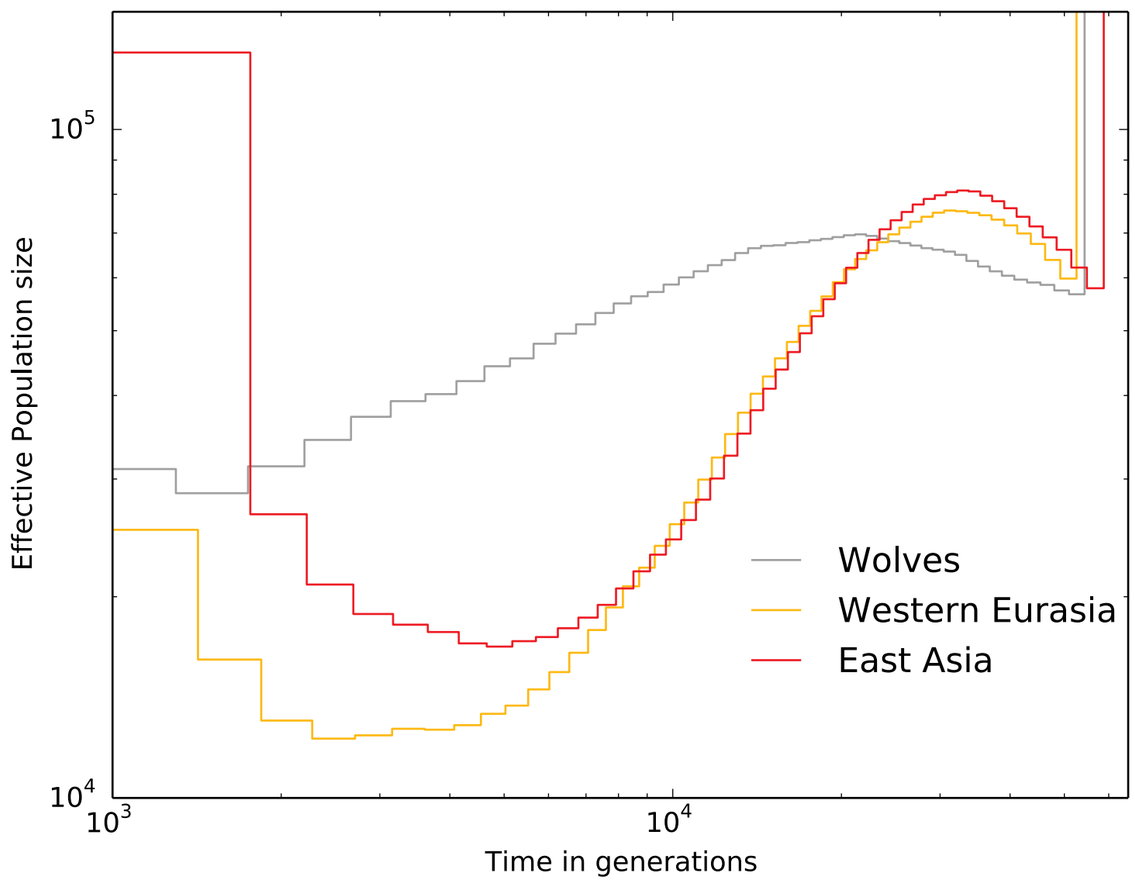
A

Modern European Breeds
Portugal village dogs (429)
Newgrange, Ireland ~ 4,800 B.P
Elkhound (12)
German Shepherd, Kunming, CzechWolf, Xolo, Lebanon Village dogs (57)
Finnish Spitz (12)
Qatar Village dog
Istrian Hound (Bosina)
Namibia Village dog
Samoyed (2)
Papua New Guinea Village dog
Eurasier (12)
Indian Village dogs (3)
Lijang Province Village dog 7
Lijang Province Village dog 8
Tawain Village dog
South China1 Village dog
Siberian Husky (2)
Greenland Sledge dogs (12)
Lijang Province Village dog 5
Indian village dog
Shar Pei (11)
Lijang Province Village dog 10
Tibetan Mastiff and Tibet Village dogs (18)
Tibet Village dog
Vietnam, Ying Province and South China2 Village dogs (13)
Lijang Province Village dogs, China (5)
Sarloos (2)
Wolves (17)
0.03

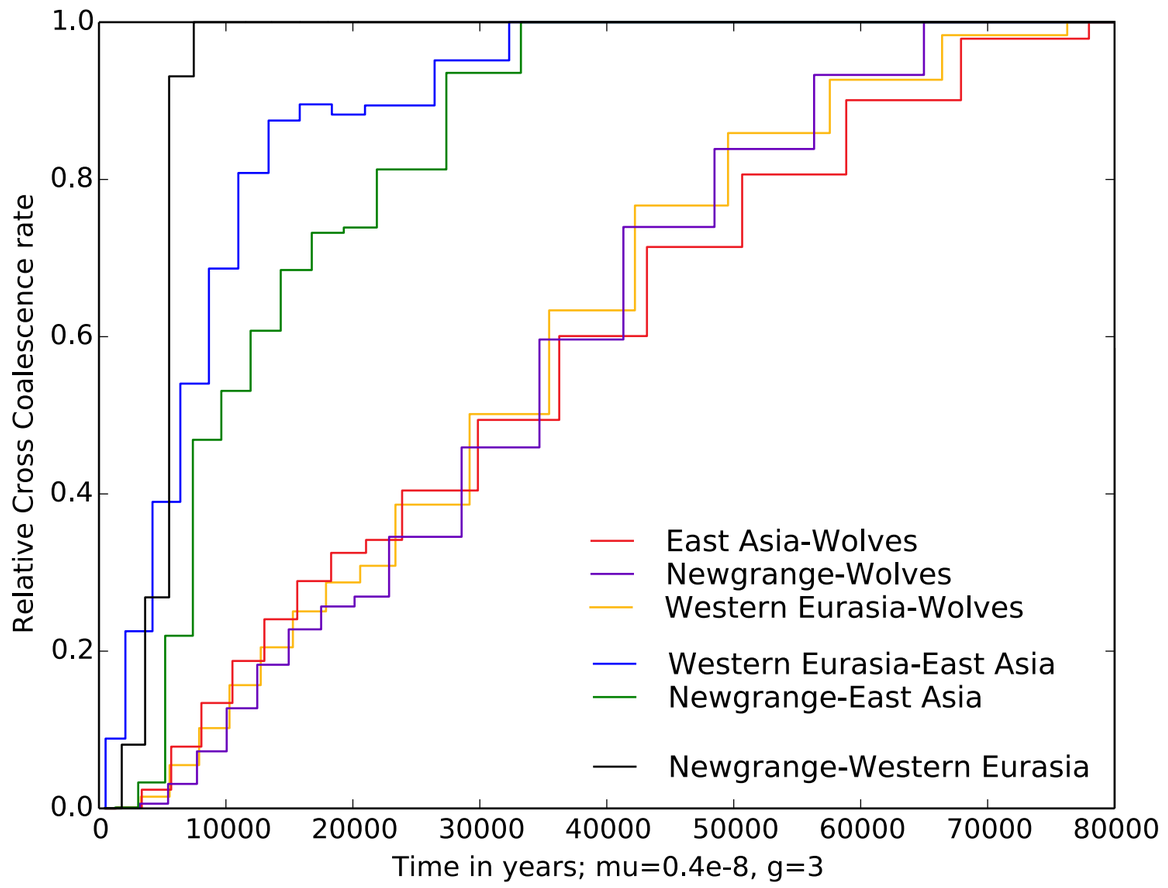
b.

Greenland Sledge Dogs
Finnish Spitz
Samoyed
Siberian Huskies
Eurasier
Newgrange dog
D
0.18
-0.1

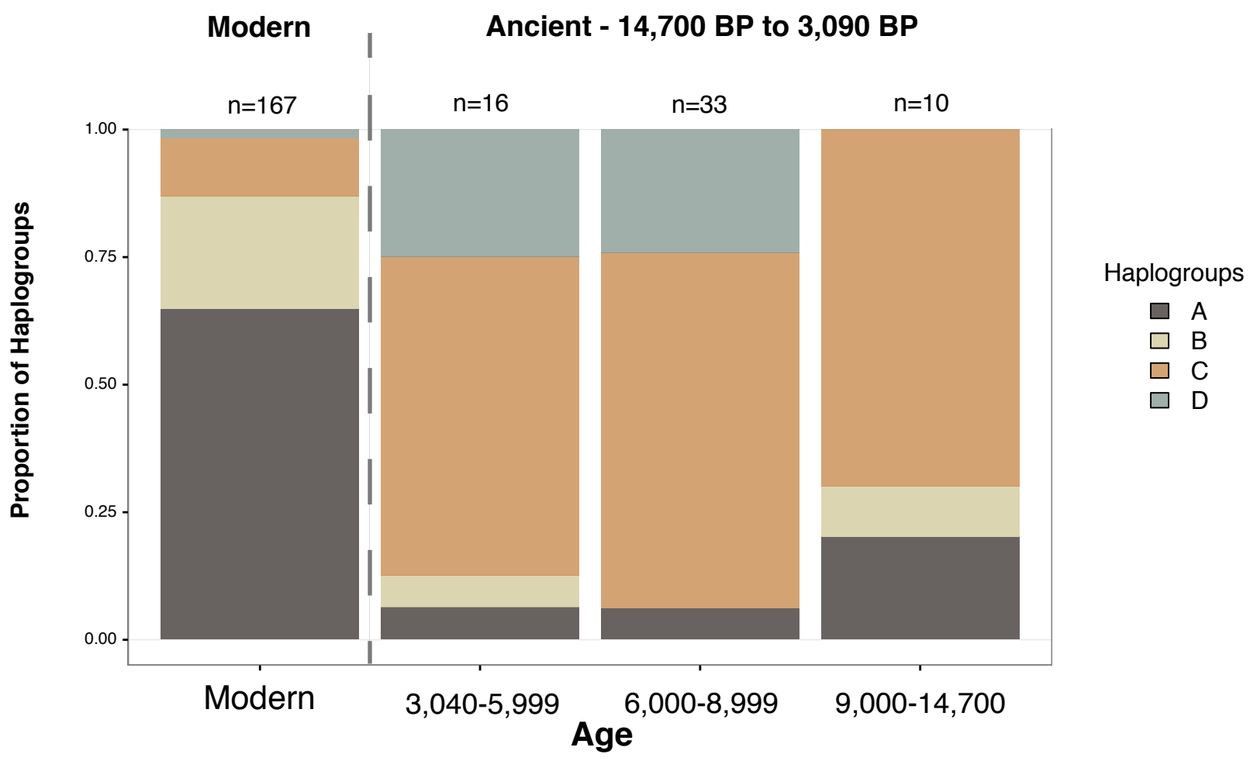
a.


Effective Population size
10^5
10^4
10^3
10^4
Time in generations
Wolves
Western Eurasia
East Asia

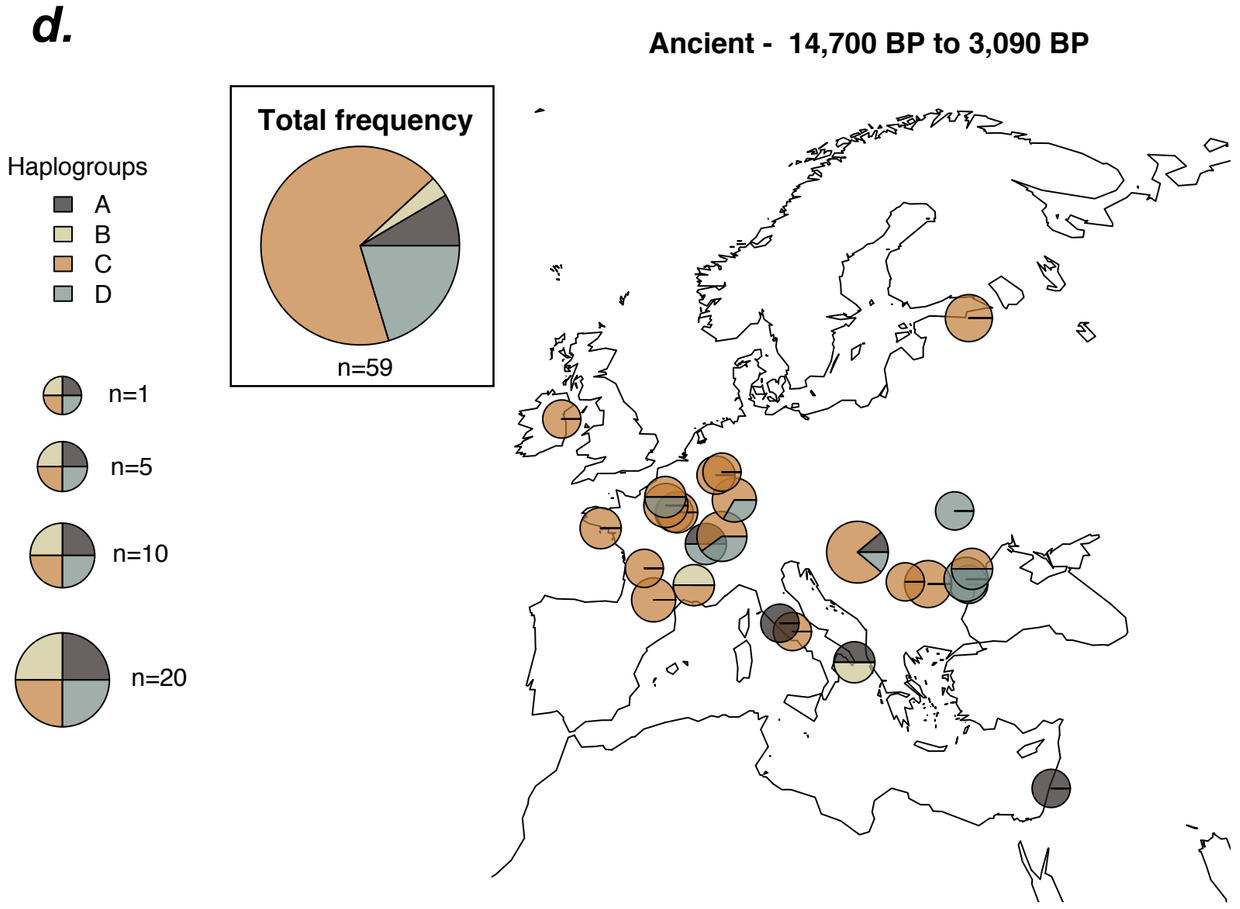
b.


Relative Cross Coalescence rate
1.0
0.8
0.6
0.4
0.2
0.0
0
10000
20000
30000
40000
50000
60000
70000
80000
Time in years; mu=0.4e-8, g=3
East Asia-Wolves
Newgrange-Wolves
Western Eurasia-Wolves
Western Eurasia-East Asia
Newgrange-East Asia
Newgrange-Western Eurasia

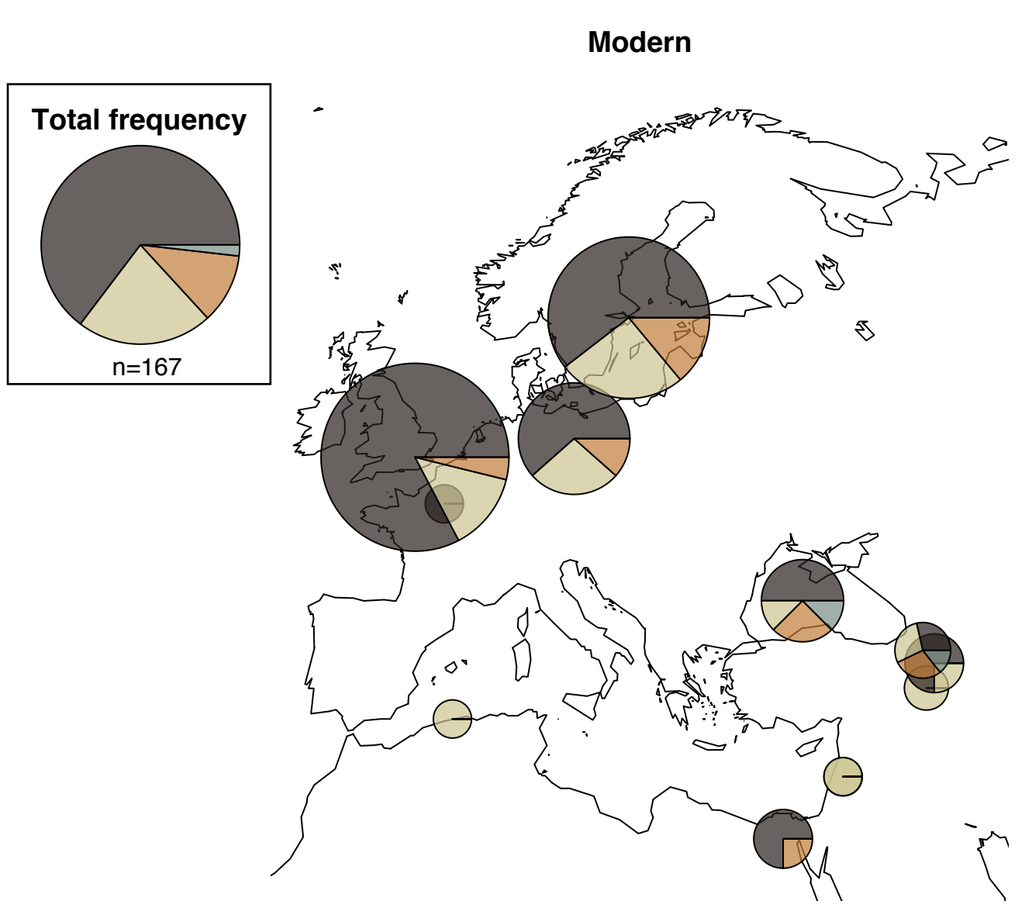
c.


Modern
Ancient - 14,700 BP to 3,090 BP
n=167
n=16
n=33
n=10
Proportion of Haplogroups
1.00
0.75
0.50
0.25
0.00
Modern
3,040-5,999
6,000-8,999
9,000-14,700
Age
Haplogroups
A
B
C
D

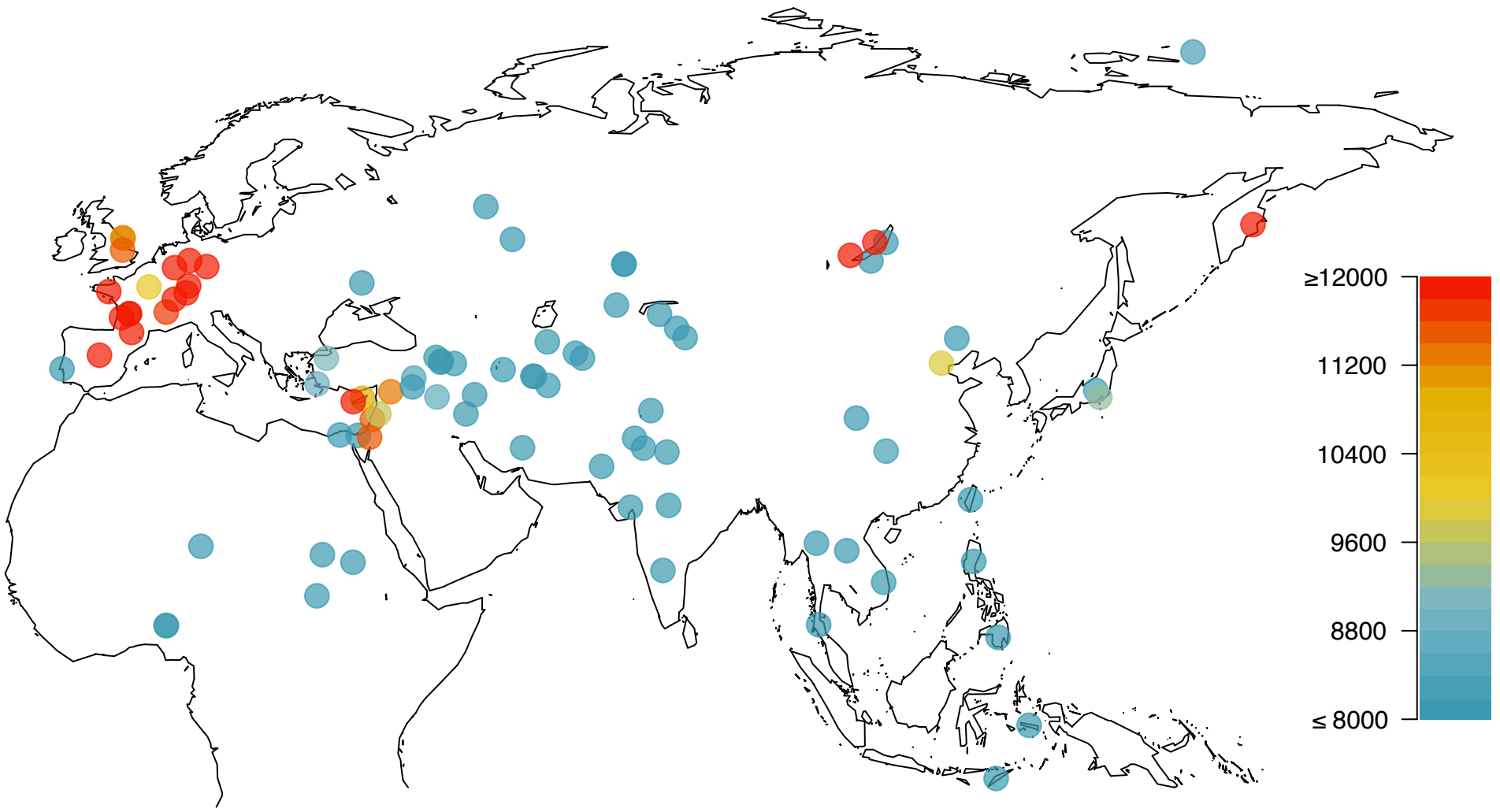
d.


Ancient - 14,700 BP to 3,090 BP
Haplogroups
A
B
C
D
Total frequency
n=59
n=1
n=5
n=10
n=20

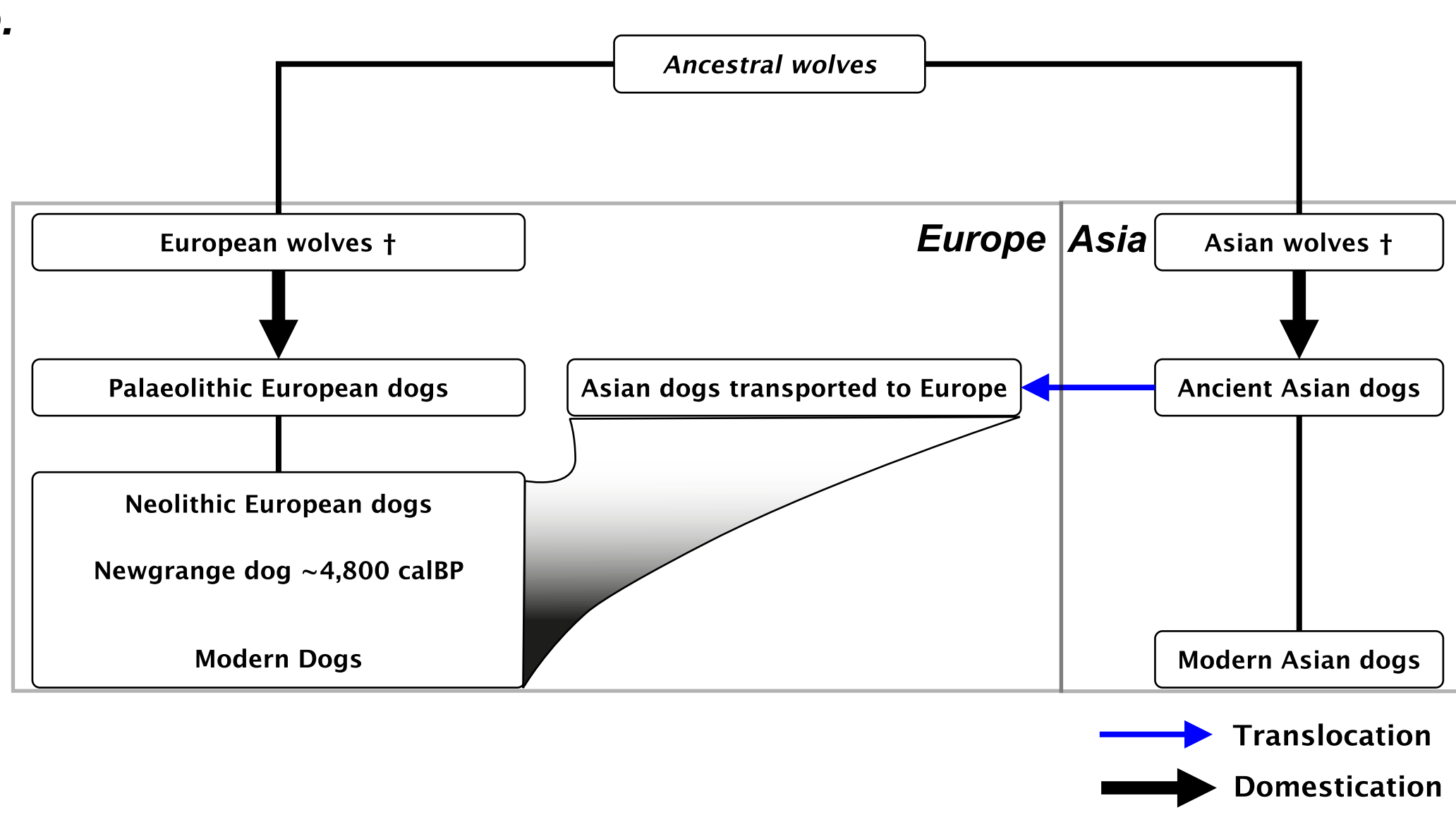
e.


Modern
Total frequency
n=167

a.


≥12000
11200
10400
9600
8800
≤ 8000

b.


Ancestral wolves
European wolves †
Europe
Asia
Asian wolves †
Palaeolithic European dogs
Asian dogs transported to Europe
Ancient Asian dogs
Neolithic European dogs
Newgrange dog ~4,800 calBP
Modern Dogs
Modern Asian dogs
Translocation
Domestication

Supplementary Materials for

Genomic and archaeological evidence suggest a dual origin of domestic dogs

Laurent A. F. Frantz*, Victoria E. Mullin, Maud Pionnier-Capitan, Ophélie Lebrasseur, Morgane Ollivier, Angela Perri, Anna Linderholm, Valeria Mattiangeli, Matthew D. Teasdale, Evangelos A. Dimopoulos, Anne Tresset, Marilyne Duffraisse, Finbar McCormick, László Bartosiewicz, Erika Gál, Éva A. Nyerges, Mikhail V. Sablin, Stéphanie Bréhard, Marjan Mashkour, Adrian Bălăşescu, Benjamin Gillet, Sandrine Hughes, Olivier Chassaing, Christophe Hitte, Jean-Denis Vigne, Keith Dobney, Catherine Hänni, Daniel G. Bradley and Greger Larson

correspondence to: Laurent A. F. Frantz (L.A.F.F) – laurent.frantz@arch.ox.ac.uk; Greger Larson (G.L.) – greger.larson@arch.ox.ac.uk; Daniel G. Bradley (D.A.G.) – dbradley@tcd.ie

This PDF file includes:

Materials and Methods

Newgrange dog archeological context and Sequencing

Archeological context

The bone element sampled here was the petrous portion of the temporal bone (**Fig. S1**). Our results show that petrous should not only be considered as the element of choice for human samples (*23*), but should also be considered for non-human species.

The dog petrous bone came from the settlement that grew up around the Irish Neolithic Passage Grave at Newgrange (*24*). Originally it was thought that the settlement dated exclusively to the early Bronze Age but re-appraisal of the pottery indicated that the presence of Late Neolithic material as well (*25*) and radiocarbon dating of the petrous bone indicates that it belongs to an earlier period (see below). Metrical analysis of the diagnostic canid remains indicated that they were of medium sized dogs and there was no evidence for the presence of wolf at the site (*24*). Moreover, besides a few isolated cases, wolves are almost never found at Irish archaeological sites (*26*). Additional genetic evidence in the following text, such as clustering analyses, functional variant assessment (starch digestion) and demographic analysis also clearly support the fact that this petrous bone was derived from a dog.

Radiocarbon age

The sample was directly radiocarbon dated at the Chrono Centre, Queens University Belfast, dating conservatively to 4900-4700 cal BP with 95.4% confidence interval calibrated using (*27*) (un-calibrated 4229±34 BP; **Fig. S2**).

DNA extraction and sequencing

Sample preparation, extraction and library preparation were performed in a dedicated ancient DNA laboratory in the Smurfit Institute of Genetics, Trinity College Dublin. Standard aDNA protocols were carried out throughout the process (*28*) including the use of extraction, library and PCR controls.

The sample was decontaminated via UV exposure, the removal of surface layers using a dental drill and further UV exposure. Using a dremmel diamond wheel a piece of bone from the densest area of the petrous was cut out and sub-sampled with the use of a mixer mill (MM 400, Retsch). All tools used in this process were cleaned with bleach, DNA-ExitusPlusTM and ethanol before UV exposure for a total of 60 minutes.

The extraction of 0.14g of bone powder was performed based on (*23*) using a two stage extraction first described by (*29*). Briefly, the bone powder was incubated at 55^{o}C for 24 hours followed by 24 hours at 37^{o}C using a thermo mixer at 700rpm with 1ml of lysis buffer (Tris HCl pH 7.4 - 20 mM; Proteinase K, recombinant, PCR Grade – 0.65 U/ml; Sarkosyl® NL 30 – 0.7 %; EDTA pH 8 – 47.5 mM). The sample was then centrifuged for 10 minutes at 13,000 rpm, supernatant removed, fresh lysis buffer added and the incubation and centrifugation steps were repeated. As contamination is more likely to be accessed during the first lysis stage (*30*) supernatant from the second extract was taken to the next stage.

The extract was purified via a two step purification; the centrifugation of the supernatant in an Amicon® Ultra-4 Centrifugal Filter Unit 30K with 3ml of 10 mM Tris-EDTA Buffer for approximately 20 minutes at 2,500 rpm or until the volume obtained within the filter was 250µl. The flow through was then discarded, fresh 10 mM Tris-EDTA buffer added and the centrifugation repeated until the volume obtained in the filter was 100µl. The final purification of the remaining 100µl was then achieved using a silica column MinElute PCR Purification Kit, (QIAGEN) with a final elution of 20µl made using a solution of TWEEN® 20 (0.05% final concentration, SigmaAldrich) added to the Elution Buffer.

A single stranded library was constructed from 30ul of DNA extract based upon the protocol of (*31*) with modifications as reported in Gamba and colleagues (2014). Briefly, Blunt end repair was performed using EBNext® End Repair Module (New England BioLabs Inc.) and during the final step of Adapter Fill in Bst activity was arrested through heat inactivation (20 min at 80°C). Indexing PCRs were performed using AccuprimeTM Pfx Supermix (Life Technology), primer IS4 (0.2 µM) and indexing primers (0.2 µM) (*31*). 3µl of library was added to a total volume of 25µl of PCR mix and amplified using 5 min at 95°C, 12 cycles of 15 sec at 95°C, 30 sec at 60°C and 30 sec at 68°C followed by a final extension of 5 min at 68°C. PCR products were purified using MinElute PCR Purification Kit, QIAGEN and assessment of amplified libraries was achieved via the Agilent 2100 Bioanalyzer using an Agilent DNA 1000 Kit and following manufacturers instructions.

The library was first screened on an Illumina MiSeqTM platform at TrinSeq (Trinity Genome Sequencing Laboratory, Trinity College Dublin, Ireland), using 50bp single-end sequencing and a PhiX control at 1%. The library was subsequently sequenced using 100 bp single-end sequencing on a total of 10 lanes of an Illumina Hiseq2000TM Platform at BGI (Beijing Genomics Institute, China

Newgrange dog data processing

Raw-read processing

To minimize the possibility of post-PCR contamination, raw reads were filtered for exact matches to the indices used in library preparation. Cutadapt (*32*) was employed to remove adapter sequences from reads of a minimum of 30bp or greater, allowing for a one base overlap between the adapter and the read.

Alignment

Sequence reads were aligned using Burrows-Wheeler Algorithm (BWA) version 0.7.5a-r405 (*33*) canFam3.1 (*34*) with the addition of the dog reference mitochondrial genome (NCBI accession number NC_002008.4), run with default parameters apart from disabling the seed option ("-l 1024") (*35*). Samtools version 0.1.19-96b5f2294a (*36*) was then utilised to remove duplicates. BAM files from different sequencing lanes were merged using the MergeSamFiles tool from Picard v1.129 (http://broadinstitute.github.io/picard/), and further duplicates removed via Samtools version 0.1.19-96b5f2294a (Li et al., 2009). MapDamage recalibration was performed

using default parameters and the --rescale option (Jonsson et al., 2013), followed by indel realignment using GATK's RealignerTargetCreator and Indel Realigner (*37*) and then filtered via Samtools for a mapping quality of 30. Genome coverage was calculated both before and after performing mapping quality (**Table S1**) via QualiMap v2.1.3 (*38*).

Read length was assessed using QualiMap v2.1.2 (*38*), and molecular damage was assessed using MapDamage2.0 default parameters (*39*) post removal of duplicates, but prior to indel realignment and mapping quality filtering. The deamination patterns seen in **Fig. S3**, an increase in C->T transitions at 5' prime ends and the complementary G -> A transitions at the 3' prime ends, and the mean read length (**Table S1**) are all indicative of postmortem damage (Gamba et al. 2014, Jonsson et al. 2013).

Sexing

Since there was a lack of Y chromosome alignment in the canFam3.1 genome build (*34*) we used a read depth based method (*17*), comparing alignment of sequencing reads post filtering for a mapping quality q30 to the X chromosome and chromosome 1. Only 1839185984 reads aligned to the X chromosome, compared with 3529477888 that aligned to chromosome 1, once this has been normalised by the length of the chromosome (1839185984/123869142) / (3529477888/122678785) = 51.6 %, similar to a previously reported male wolf genome (*17*). This evidence suggests that the Newgrange dog was male.

Publically available data and SNP calling

We made use of 124 genomes from various dog breeds and village dogs sampled across the world. We downloaded 77 BAM files from the *DoGSD* database (*40*). We also downloaded raw reads for an additional 51 dogs and an Andean fox (used as an outgroup) from NCBI Short read archive (*41*). The raw reads from these 51 samples were aligned to the canFam3.1 reference genome using bwa mem (*33*). We computed depth of coverage (DoC) for each samples using bedtools (*42*). We only kept genomes with a minimum DoC of 10x to ensure reliable SNP calling (*43*). This resulted in 80 high quality data from breeds, village dogs and two wolves (**Table S2**). We also made use of data from 525 dogs and wolves genotyped on the CanineHD 170K SNP array (*44*).

SNP calling

We used samtools 'mpileup' (0.1.19; (*36*)) to call SNPs with default settings. Pileup files were further filtered, for each sample, using the following criteria:

- Minimum DoC >= 6
- Excluded all sites in region of high DoC (top 5%)
- Excluded all sites within 3bp of an indel
- Minimum Phred >= 20
- Minimum fraction of reads supporting heterozygous >= 0.3
- For the Newgrange dog, we also discarded the first and last 8bp of each read for SNP calling, to avoid incorporating errors from deaminated sites (see above).

For the Newgrange dog, we ran this procedure on both re-calibrated (with MapDamage; see above) and raw BAM files. Total number of SNPs per sample is

reported in **Table S3**. SNPs were merged using bedtools. Ascertainment was done without the outgroup, all non bi-allelic markers were also removed as well as markers with more than 20% missing data. This resulted in 19,388,128 high quality SNPs.

Lastly we repeated the same procedure solely on sites that were called on the 170K array (see above). The SNPs on this array are mapped on canFam2.0. Thus, we used liftOver (*45*) to map these onto canFam3.1. Merging was done with plink (*46*). We also ascertained SNPs solely in the genome data base (80 individuals) using plink. To do so we ran plink with the following options:

--maf 0.05
--geno 0.1
--bp-space 5000

This resulted in 348,927 variants. To avoid any problem arising from DNA damage in the Newgrange dog, we also created a set of SNPs using only transversions using the same Plink command. This resulted in 269,512 SNPs.

Ancestry analyses

Newgrange dog ancestry

We combined the 81 individuals (including the Andean fox as an outgroup) above with the 525 samples available on the 170K HD array, totalling 605 dogs and an outgroup. We used plink to compute an Identity By State (IBS) matrix. This matrix was used to build a neighbour joining tree (NJ) using the R package "ape" ((*47*); **Fig. 1a**). To compute support for each node we bootstrapped our ped file 100 times and re-computed a NJ tree each time. The same analysis was repeated using solely the genome samples (with variant ascertained only in those samples) to ensure the HD array ascertainment did not influence our basic clustering (**Fig. S4**). The result of these clustering analyses shows a strong affinity between modern European breeds and the ancient Irish dog. In **Figure S4**, the Newgrange dog clusters with Lebanon village dogs, Portuguese village dogs as well as Xoloitzcuintle (Mexican hairless dog), Labradors, Kunming and German shepherd breed dogs. Xoloitzcuintle and Kunming dogs are not from Western Eurasia (Mexican and Chinese respectively); however both of these breeds have been shown to be mainly derived from European dogs (*8*, *18*) further strengthening the tight relationship between the Newgrange dog and modern European dogs.

To further test this finding we ran a PCA on the 170K SNP with these 605 samples using *smartpca* (*28;* **Fig. S5**), as well as a *TreeMix* analysis ((*49*); **Fig. S6**). For the *TreeMix* analysis, we first subsampled European breeds for conciseness (see **Fig. S6**) we then fitted a simple model without migration edges with bins of 100 SNPs. In addition, we added 1-3 admixture edges to investigate the pattern of admixture among dog populations (**Fig. S6**). We also computed D-statistics using qpDstat (*50*, *51*). For each pair of sample (A, B) we computed D(Outgroup, Newgrange; A, B) (**Fig. S7**). Standard error for this statistics was obtained by performing a weighted block jackknife over 5Mb blocks. On **Figure S7**, the y-axis represents B in D(Outgroup, Newgrange; A, B). A positive values, therefore, implies that the Newgrange share more derived allele the population on the y axis, while negative values imply that the Newgrange dog is closer to

an other population (A). D values where B=European are mainly positive and negative where A=Asian. This demonstrates that the Newgrange dog shared more derived alleles with modern European breeds than any other population tested in this study.

We also ran a PCA solely on the genome-wide data set (see above), with and without transitions (**Fig. S8**; **Fig. S9**). Removing transitions had no effect on this analysis. Interestingly, the same pattern arises, with the Newgrange dog being closely related to Portuguese and non-German Sheperd European breeds on PC1 (**Fig. S8**; **Fig. S9**). Moreover, PC1 clearly differentiate East Asian and Western Eurasian samples. However, most of the PC1 axis highlights the high degree differentiation of German shepherd (GS; and Kunming dogs, which are mainly derived from German shepherd). This is not surprising given their high degree of inbreeding found in this breed (see RoH analysis below). Moreover, this is likely the result of ascertainment bias given the high proportion of GS in this sample set (9/80).

Interestingly, PC2 seems to be differentiating modern dogs from the Newgrange dogs (**Fig. S8**). This result holds when only using transversions (**Fig. S9**). Moreover, these figures contain with and without re-scaled genotype (see above) for the Newgrange dog. Both re-scaled and non-rescaled (via mapDamage; see above) are plotted exactly at the same position and are thus indistinguishable on the plot. Together these analyses suggest that this result is not due to ancient DNA damage. Thus, while this component explains much less of the variance (PC1~11%; PC2~4%), it suggests that some of the ancestry found in this ancient Irish dog may not be reflected in modern Western Eurasian populations (see "*The Newgrange dog may retain some degree of ancestry from an ancient (extinct) European population***"**).

Defining "core" groups

The tree in **Figure 1a** highlights a deep split between Western Eurasian and East Asian dogs. We used D-statistics to assess the robustness of this split. To do so we first defined "core" groups. We used all modern breeds as well as the Portugal village dogs that formed a strong cluster in **Figure 1a** (429 individuals; bootstrap value = 100) to define the Western Eurasian "core" group. For the East Asian grouping we used all samples that had a good support (>90) and clear East Asian origin (**Fig. 1a**). This includes Sharpei, Village dogs from China, Tibet and Vietnam as well as Tibetan mastiffs.

Ancestry of non-core populations

In order to assess the affinity of every population in the dataset to each of the two cores, we computed values for all possible pairs of D-statistics including both Western Eurasian and East Asian core populations such as D (Outgroup, X; East Asia, Western Eurasian). Highly negative and positive D values support the East Asian and Western Eurasian affinities respectively (**Fig. S10**). This degree of derived allele sharing is also plotted in **Figure 1b**. Thus, while we find that many "basal" breeds (*1*) (*e.g.* Eurasia, Siberian Husky or Greenland Sledge dog) as well as multiple Asian samples (*e.g.* Papuan New Guinean village dogs) show affinity toward the Western Eurasian core, their D value are close to 0 suggesting mixed ancestry (**Fig. S10**). This is further supported by

admixture edges obtained from TreeMix, in which both Papua New Guinean (PNG) and Indian dogs show admixture with Asian core (**Fig. S6**).

Mixed ancestry of non-core populations

We then assessed whether each population possessed mixed ancestry. To do so, we computed the significance (Z) of every possible pair of D(Outgroup, East Asian; X, Western Eurasian). We found that all core Western European populations as well as most other European populations show little to no significant evidence for admixture from the East Asian core (**Fig. S11**). However, populations of basal breeds such as Eurasier, Greenland Sledge dogs and Siberian Huskies show significant traces of admixture (**Fig. S11**). Moreover, East Asian populations, that clustered with Western Eurasian dogs (*e.g.* Papua New Guinean village dog; **Fig. 1a**) also show clear signs of admixture from the East Asian core samples (**Fig. S11**) as shown in our TreeMix analysis (**Fig. S6**).

Wolf dog admixture

Here we assessed every pair of D(Outgroup, Wolf; A, B). As expected, we found that the Czech wolfdog and Sarloos are admixed with wolves (**Fig. S6**; **Fig. S12**). Interestingly, the Newgrange and the Lebanon dog also show signs of admixture when compared to other European populations (**Fig. S12**). Moreover we found that wolves always share more derived allele with Western Eurasian that with East Asian (**Fig. S13**), suggesting admixture from wolves into European populations or *vice-versa*.

To further test this hypothesis we fitted two different models with *ADMIXGRAPH* (*51*) to a reduced data-set (**Fig. S14**). We found that a model without admixture from wolves resulted in a fit with two f_4 outliers: f_4(Andean_Fox; Wolf; Newgrange, Ying Village Dog) $Z > 3$ and f_4(Andean_Fox; Wolf; Portugal Village Dog, Ying Village Dog) $Z > 3$ (**Fig. S14**). Including admixture from wolves in the model, however, left no outliers (**Fig. S14**). The same pattern (excess of shared derived allele between wolves and European samples) using D-statistics based on whole genome SNPs (**Table S4**). We note that such a signal could be the result of admixture from modern wolf into dogs or *vice-versa* (*6*).

Multiple sequentially Markovian coalescence (*MSMC*)

We used the Multiple sequentially Markovian coalescence (*MSMC*; (*14*)) to reconstruct the population size of Wolves, East Asian, Western Eurasian populations through time. We used BEAGLE 4.1 with default settings (*52*) to phase the data. For better resolution we used the recombination map of the dog (downloaded from the Auton lab at http://autonlab.einstein.yu.edu/ dog_recomb/) and phased all individuals in the same run. *MSMC* requires a good depth of coverage to ensure that heterozygous positions are called (*14*). Therefore, we used only high coverage data: two Chinese village dogs from the Ying province (~17x) two Portuguese village dogs (~17x) and two wolves from Russia (~12x; **Table S2**). We also generated a mask for mapability using the SNPable toolkit (http://lh3lh3.users.sourceforge.net/snpable.shtml - with k=35 and r=0.5) to exclude non-unique regions in the dog genome. We first ran *MSMC* on each genome separately (*PSMC'*; **Fig. S15**). Thereafter we ran *MSMC* on each pair of sample (3 populations: Chinese and Portuguese village dogs and wolves) separately (**Fig. 2a**).

Lastly, we also computed cross-coalescence rate (CCR; (*14*)) between every pair of populations (**Fig. 2b**).

Using *MSMC* to estimate the canid mutation rate

MSMC requires input of mutation rate per generation (*mu*) as well as generation time (*g*) for estimating time (in years). We used a similar approach as (*17*) taking advantage of the radiocarbon age of the Newgrange dog to compute the mutation rate. More precisely, we computed the CCR between the Newgrange dog and the Portuguese village dogs and translated the estimated divergence time from generations into years using various mutation rates (**Fig. S16**) and a generation time of 3 years as in other studies (*6*, *9*, *17*, *34*, *53*). Our results are consistent with a mutation rate between 0.3 and $0.45e^{-8}$ mutation per year, similar to previous estimates using ancient DNA (*17*). The resulting inferred time of divergence between each population is displayed in **Figure S17**. This figure suggest that the time of divergence between East and Western Eurasian dogs is younger than the earliest appearance of archeological dog remains in Europe.

mtDNA sequencing and analysis

Modern mtDNA

The d-loop sequences from 2,586 dogs, from multiple studies (*19*, *54–64*), were downloaded from genebank (**Table S5**).

Ancient mtDNA.

Multiple ancient d-loop / full mitochondrial genomes from several sources (*7*, *65–67*) were also obtained from Genbank (**Table S5**). These represent 7 out of 59 ancient sequences from European dogs used in this study. We also generated 66 ancient d-loop sequences from 26 sites at the French National Platform of Paleogenetics, France (PALGENE, CNRS, ENS de Lyon) (*68*)(**Table S5**). In addition, 34 ancient d-loop sequences were generated at the Durham Evolution and Ancient DNA (DEAD) lab at Durham University, UK (*69*).

Ancient DNA extraction and sequencing at ENS, France

The external bone surface was scratched with a sterile scalpel and the remaining cleaned part was reduced to powder with a sterile hammer. The powder (150–300 mg) was then digested for 18 hours at 55°C with agitation in 4.7 ml of buffer (0.5 M EDTA, pH=8.0), 50 µl of proteinase K (1 mg/mL) and 250 µl of 0.5% N-lauryl-sarcosine. A silica-based method modified from (*70*) was used to retrieve DNA. Mock extractions were performed in order to rule out contamination from reagents. In addition, cross-contamination was monitored using samples from other species (sturgeons and sheep) together with dog samples in every single DNA extraction session.

Ancient amplification and sequencing pre-amplification procedures were performed in the PALGENE platform, while post-amplification analyses were conducted in a post-PCR molecular laboratory located in a different building. Aerosol controls, consisting of a tube left open throughout manipulation, were included in all PCR assays. Additional PCR controls, including negative PCR-mix controls and extraction blanks, were also added for every amplification attempt.

Three overlapping fragments were targeted in order to cover the 139-bp fragment in the mitochondrial HVRI, using the following primer sets: “Canis F1” 5’GTGCTATGTCAGTATCTCC3’ – “Canis R3” 5’ATRTAATATYATGTACATGC3’ (64 bp, including primers); “Canis F2” 5’TGGTTTGYCCCATGCATA3’ – “Canis R4” 5’TGATTAAGCCCTTATTGGA3’ (97 bp, including primers), and; the pair of primers DL1 –DL3 (139 bp, including primers) defined in Leonard et al. (2002). All amplifications were carried out in 25 µL containing 1-5 µL of DNA extract, 1 mg/mL BSA (Roche, 20 mg/mL), 250 µM of each dNTP, 0.5 25µM of each primer, 2.5 units of Perkin Elmer Gold Taq polymerase (Applied Biosystems), 2.5 µL of 10X buffer, 2 mM of MgCl2. Cycling conditions were as follows: one activation step at 94°C for 5 min followed by 60 cycles of denaturation at 94°C for 30 s, annealing at 50-55 °C [depending on primers] for 30 s, elongation at 72 °C for 45 s, and a final extension at 72 °C for 10 min.

Positive PCR amplicons were cloned using Topo TA Cloning for Sequencing kit (Invitrogen). Clones were picked and tested for inserts by PCR with the ready-to-use Mastermix (Eppendorf, Hamburg, Germany) using the following cycling conditions: 40 cycles at 94°C for 30 s, 55°C for 30 s, and 72°C for 45 s; and at 72°C for 5 min. Products of the expected size were sequenced on both strands by Genome Express, Grenoble. The sequence of a given PCR product was determined as the consensus of the sequences of a minimum of 24 clones (max. 48 clones). Each fragment was amplified at least twice in two independent PCR sessions and the final sequence of a given fragment was determined as the consensus of all clone sequences from at least two PCR products. Further information about the archaeological context and the dating can be obtained from Pionnier-Capitan (2010).

Ancient DNA extraction and sequencing in the DEAD Lab at Durham University, UK.

To prevent sample contamination and ensure the generation of authentic ancient data, strict DNA procedures were followed based on (*71*, *72*). An electric hand-drill was used to remove approximately 1mm of the exterior surface of the bone, after which each sample was ground to fine powder in a sterilised stainless steel canister using a microdismembrator (Sartorius). Samples were incubated at 50°C overnight in a solution containing 1.7ml of 0.5M EDTA (pH8), 0.2ml of 10mM TRIS-HCl (pH8), 0.1ml of SDS (1%w/v) and 20µl of Proteinase K (20mg/ml). Purification was undertaken using the QIAquick PCR purification Kit (QIAGEN Ltd, UK) following the manufacturer’s instructions.

The forward primer ‘Canis F2’ (5’-TGGTTTGYCCCATGCATA-3’) and the reverse primer ‘Canis R4’ (5’-TGATTAAGCCCTTATTGGA-3’) (see above) were used to amplify a 97bp fragment (excluding primers). This fragment corresponded to the positions 15,572 – 15,669 on the complete dog mitochondrial DNA genome (*73*). DNA extractions of longer fragments were attempted on samples from various regions but the success rate was extremely low. The highest success rate was obtained through the shorter 97bp fragment. The PCR amplification was performed in 25µl containing 2µl

extract, 0.76x PCR Gold Buffer, 1.89mM MgCl2, 1.04U Taq, 0.18mM dNTP and 0.75µM of each primer. 2.5µl of BSA (25mg/ml) was added. Thermo-cycling conditions comprised of a 3-minute denaturation step at 94°C, followed by 50 cycles of 45s denaturation at 94°C, 45s annealing at 55°C, 45s at 72°C, and a 10-minute final extension step at 72°C. If PCR amplification failed, the number of cycles was increased to 70.

The PCR amplifications were visualised on a 1.5% agarose gel. 33 samples yielded positive results. No modern contamination was identified for the DNA extraction and the PCR blanks. Sequencing was performed on a 48-capillary 3730 DNA Analyser in a DNA laboratory located in a physically separate building (Genomics, School of Biological and Biomedical Sciences at Durham University). The sequencing primers were identical to the primers used for DNA amplification. Sequencing was undertaken on both strands. Further information about the archaeological context and dating of each of the samples is listed in (*69*).

mtDNA haplogroup assignment

We assigned each d-loop sequences to one of the four major mtDNA haplogroups, A, B, C and D (*19*) using *DomeTree* (*74*). However, *DomeTree's* database solely uses modern sequences and thus may lack some variation only found in ancient sequences. *DomeTree* also provides a database with diagnostic mutation for haplogroup assignment. We therefore used this database to assess the reliability of the haplogroup assignment provided by *DomeTree*. To do so, we first aligned these sequences using muscle (*75*) and for each sequence we:

1) Computed the proportion of mutations to the reference that were diagnostic for each haplogroup proposed by *DomeTree*.
2) Assigned each sequence to a clade if:
    a. DomeTree reported only one clade
    b. If more than 70% of the mutations (to the reference) are single diagnostic mutations (only found in a single haplogroup).

In all, 207 out of 2,744 (<20%) of the d-loop sequences in our dataset could not be assigned to a haplogroup using this method. Of these, 114 were ancient samples. **Figure S18** shows the proportion of variants for each ancient and modern sequence that was recognized by *DomeTree* (existing in its database). This figure highlights a clear bias toward modern sequences, suggesting that many of the ancient mutations are absent from modern populations. In order to classify the sequences that did not pass our filters, we computed all combinations of nucleotide distance between unclassified sequences and sequences that passed the above filters. We assigned each sequence to a clade if:

1) The difference between the distance to the closest haplogroup and the second closest haplogroup was greater than 2
2) The difference was at least 1 and the sequence shared a common indel with a sequence that was assigned to a haplogroup in the previous filtering (see above).

Note that transversions were counted as a distance of 2. This allowed us to classify all of the remaining samples to a clade except a final 53 sequences. Lastly, we built a minimum spanning network using PopART (*76*) with all 2,739 sequences to classify the remaining 53 sequences. From the network we assigned each sequence to the closest clade. If two clades were equally close the sequence was kept unassigned. This approach allowed us to assign 41 sequences out the 53 remaining sequences. The 12 non-assigned sequences were discarded from further analysis. Contextual information including accession, age, geographic coordinates and haplogroup used in this study (2,744) are available in **Table S5**.

Drift simulations

Our mtDNA analysis revealed the existence of a turnover in haplogroup frequency. We attribute this turnover to a migration of dogs into Europe with haplgroup A that replaced the ancestral haplogroup C (see main text). However, this pattern could be explained by drift alone (no additional influx of haplogroup A). We used a simulation to investigate this issue.

We simulated genealogies using *ms* (*77*). To obtain a reliable demographic history we used our *MSMC* estimates (**Fig. 2a**). We used $Ne_{mtDNA}=1/4*(Ne_{nucDNA})$ to mimic the effective population size at the mtDNA (matrineal lineage, haploid). We simulated genealogies in a population, with two clades, one clade at a starting frequency of 0.1 and another at 0.9. We then computed the probability that the clade starting at 0.1 reached a frequency of 0.66 and above as in our mtDNA for haplogroup A. To do so, we simulated 167 sequences and computed frequency of the minor haplogroup (starting at freq=0.1) 100000 times for 1000 and 5000 generations (roughly corresponding to 3000 and 15000 years ago – **Fig. 2c**). For both cases (1 and 5K generations), we computed the probability that the starting minor clade reached any given frequency (or higher) at t=0 under our demographic model by dividing the number of times we observed a frequency equal or higher by the number of replicate simulations (100,000).

The following *ms* command were used:

15,000 years (5000 generations):

ms 170 100000 -T -I 4 1 1 1 167 -t 0.00016 -es 0.125 4 0.9 -ej 0.2 2 4 -ej 0.2 5 3 -ej 1 4 3 -ej 1.5 3 1 -en 0.00248115625 4 0.629566 -en 0.0407404375 4 0.402687 -en 0.051380375 4 0.326363 -en 0.0622155625 4 0.306688 -en 0.073253125 4 0.310168 -en 0.08450125 4 0.31726 -en 0.0959675 4 0.316277 -en 0.10766125 4 0.321211 -en 0.11959125 4 0.334041 -en 0.131766875 4 0.343619 -en 0.14419875 4 0.363365 -en 0.156898125 4 0.385689

3,000 years (1000 generations):

ms 170 100000 -T -I 4 1 1 1 167 -t 0.00016 -es 0.025 4 0.9 -ej 0.2 2 4 -ej 0.2 5 3 -ej 1 4 3 -ej 1.5 3 1 -en 0.00248115625 4 0.629566 -en 0.0407404375 4 0.402687 -en 0.051380375 4 0.326363 -en 0.0622155625 4 0.306688 -en 0.073253125 4 0.310168 -en 0.08450125 4 0.31726 -en 0.0959675 4 0.316277 -en 0.10766125 4 0.321211 -en

0.11959125 4 0.334041 -en 0.131766875 4 0.343619 -en 0.14419875 4 0.363365 -en 0.156898125 4 0.385689

Figure S19 shows the result of this analysis. We found that the probability of a clade to reach 0.66 (starting at 0.1) after 3,000 and 5,000 generations under our demographic model is 0.00535 and 0.031 respectively. This analysis thus clearly demonstrates that a model involving drift alone is unlikely to explain our results.

The Newgrange dog may retain some degree of ancestry from an ancient (extinct) European population

MSMC and simulations

Interestingly, the time of divergence between East Asian and the Newgrange dog seems on average older than the time of divergence between Western European and East Asian (**Fig. 2b**; **Fig. S17**). This is clear from the faster decay of CCR between Newgrange and Asian than between Western Eurasia and Asian (**Fig. 2b**; **Fig. S17**). Here we assessed whether this signal signal could stem from the fact that the Newgrange dog retained a degree (*f*) of ancestry from an ancient, now extinct, population of dogs that inhabited Europe prior to its replacement.

To do so we simulated data under this model and computed CCR with MSMC. In this model we assumed that modern dogs diverged from the now extinct population from Europe 10,000 generations ago (30,000 years), modern European and the Newgrange dog diverged 4,800 years ago (1,600 generations) and that the European and Asian core diverged 8,000 years ago (~2,600 generations; **Fig. S20a**). We simulated 30Mbp fragments with macs (*64*) using the following command (*f*=0.1 and 0.2):

macs 8 30000000 -t 0.0004 -r 0.00016 -I 4 2 2 2 2 -es 0.04 2 1-f -ej 0.0401 5 4 -ej 0.0402 3 2 -ej 0.066 2 1 -ej 0.3 4 1 -eN 0 1 2

Cross-coalescence rates (CCR) were then computed with *MSMC*. The result of this analysis shows that CCR can be "delayed" under this model similarly to what is observed between the East Asian lineage and the Newgrange dog (**Fig. S20b; Fig. 2b**).

The difference in CCR observed between Newgrange-Asian and European-Asian in Fig. 2b could also be explained by secondary gene flow from Asian dogs into modern European dogs (**Fig. S20c**). We also tested this model using simulations. For consistency, this model was parameterised the same way as above. We used the following macs command (again with *f*=0.1 and 0.2):

macs 8 30000000 -t 0.0004 -r 0.00016 -I 4 2 2 2 2 -es 0.04 2 1-f -ej 0.0401 5 1 -ej 0.0402 3 2 -ej 0.066 2 1 -ej 0.3 4 1 -eN 0 1 2

We find that this scenario also affects the CCR, making the time of divergence between European-Asian, artificially younger (**Fig. S20d**). However, we note that the secondary gene-flow model has a substantially smaller effect on the CCR than the previous simulated scenario (**Fig. S20a**).

We also tested whether the CCR pattern observed between the Newgrange and Asian dogs (Fig 2b) could be the result of admixture from modern wolves. To test this hypothesis we computed Newgrange-Asian CCR without "wolf like" segments in the genome. To identify these segments, we used *BEAGLE 4.1* (*52*) to detect IBD in the genome of the Newgrange. We computed IBD scores between the Newgrange genome and Asian core samples (Ying Village dog) as well as between the Newgrange genome and wolf genomes. We then extracted segments that were closer to the wolf than the Ying dogs (LOD>3). This resulted in 12 Mbp of "wolf like segments" (with LOD>3) in the Newgrange genome. We then masked these segments and re-estimated CCR with *MSMC*. This analysis had no effect on the divergence time estimates between Newgrange and Asian core (25% CCR: 8404y and 8500y for with and without wolf segments; 50% CCR: 10773y 10749y with and without wolf segments; 75% CCR: 20565y and 20519y with and without wolf segments). This result shows that the observed Newgrange-Asian CCR pattern cannot be attributed to gene flow from modern wolves.

PCA

On **Figure S8** and **Figure S9,** the Newgrange dog ancestry explains a large portion of PC2. As we demonstrated before, this is unlikely to be due to ancient DNA damage since this result holds with and without transitions (**Fig. S9**). Moreover, we show that re-calibrating quality scores using mapDamage has an effect on the number of heterozygous sites called at transitions sites (see RoH analysis below), demonstrating that re-calibration has an effect on DNA damage correction. Thus if any of this PC2 was explained by DNA damage, we would not expect re-calibrated and raw data to be plotted exactly at the same position (**Fig. S8**).

To further explore whether there is missing ancestry in the Newgrange sample not represented in modern population, we ran a PCA solely on non-East Asian genome-wide samples (solely on transvertions ascertained in non-East Asia; 242,745 SNPs). This result shows a similar pattern, with PC1 and PC2 explaining ~11% and ~4% of the variance respectively (**Fig. S20**). Here, PC1 discriminated between German shepherds (blue dots on the far right) and other dogs, while PC2 seems to follow an East to West gradient from India to Europe (Indian village dog to Portuguese village dogs and Modern European breeds such as Labrador).

In addition, the second half of the PC2 is explained by a differentiation between modern Europeans and the Newgrange dog as in **Figure S9**. To explore whether this signal can also be detected only with modern non-East Asian populations (without the Newgrange dog) we computed eigenvectors solely with modern samples and projected the Newgrange dog (**Fig. S22**). Interestingly, on both PC1 and PC2, the Newgrange is projected close to individuals that are also the closest on **Figure S21,** thus suggesting that part of this differentiation on PC2 can be explained by other samples than the Newgrange ancestry.

Admixture

We used ADMIXTURE (*79*) to assess the population structure with non-East Asian samples. We used the same 242,745 transversions as for the PCA (see above). We used a 5-fold cross-validation procedure to test the fit of K=1 to K=7. **Figure S23a** shows that the best K is 2. Here we report K2-4. In **Figure S23b,** we can see two clear populations within these samples, one including Newgrange dogs and non-German Shepherds, and another including solely German Shepherds. This pattern is similar as in the PCA above where there are two clear clusters of dogs in Western Eurasia.

Unsurprisingly, the next population to split with K=3 are the Indian village dogs most likely due to their ancestry being a composite of Western and East Asian populations (**Fig. S23c**). Interestingly, some populations such as Lebanon, Istria Hound, Namibian and Qatari also share some ancestry with Indian dogs. K=4 induced another separation among Western Eurasia dogs, separating Istria Hound, Namibian and Qatari from the Newgrange, Xolo and Labradors. All together these results point toward the same conclusion as the PCA above, demonstrating that the Western Eurasian lineage can be divided in multiple subpopulations, with some samples appearing closer to the Newgrange dog than others.

ADMIXGRAPH

The MSMC and PCA analyses suggest that the Newgrange dog may retain some ancestry that is not found in modern Asian dogs. This pattern could be explained by a partial replacement of ancient European dogs by dogs from Asia. Under this hypothesis, we would expect modern European/Asian dogs to share more derived allele than Newgrange/Asian dogs. To test this hypothesis we used *ADMIXGRAPH* (see above) and fitted a model in which Asian dogs admixed with European dogs (Portuguese village dogs) in two waves (**Fig. S24**). This model thus implies that the Portuguese dogs, rather than the Newgrange dog, are more closely related to the Asian dogs (**Fig. S24**). This model left no outliers suggesting a good fit. However, we note that this model has one additional parameter relative to our previous model (see **Fig. S14**) which also did not possess an outlier. Thus, this pattern could be due to over fitting and as a result, we cannot discriminate between these two models, as these are equally consistent with the data. Nevertheless, given the PCA and MSMC result (see above) this model appears to be more consistent with the data.

Archeological data

We compiled archeological information over 90 dog specimens from all over Eurasia as well as North Africa (**Table S7**). We first used all the information found in Larson et al. 2012 (*1*) for Eurasia and Africa. We updated this information with recently published data (Table S7) and with further sites in the Central parts of Eurasia (*7*, *80–107*). We focused our literature survey on the earliest possible sites to cover most of Eurasia. These sites comprise all the earliest (non-controversial; see below) dog remains found in Western, Eastern and Central Eurasia as well as Africa (Fig. 3a).

We are aware of proposed transitional (wolf-to-dog) specimens or "proto-dogs" from earlier European and southern Siberian Paleolithic sites dating from c. 36,500-17,000 calBP. These include specimens from Goyet Cave (*83*), Razboinichya Cave (*108*),

Kostenki 8 (*109*), Predmostí (*2*, *109*), Chauvet Cave (based on footprints)(*110*), and Eliseevichi I (*111*). However, the taxonomic identification (as dogs or wolves) of these early specimens is highly controversial (*112–115*); these were therefore excluded from our analysis (marked with a * in **Table S7**).

Heterozygosity and Runs of Homozygosity

Here we want to assess the recent demographic history of the Newgrange dog and compare it to modern dogs to assess the extent of artificial selection that was applied on this ancient sample. To do so, we first computed genome-wide heterozygosity. Estimate of herozygosity could be biased due to damage at transitions (t_i) sites in the ancient sample. To correct for this we first estimated the ratio of transversion (t_v) / transitions (t_i) at homozygous non-reference SNPs and heterozygous SNPs in all samples.

Figure S25 shows that tv/ti at homozygous sites is similar in our ancient sample. However, this ratio shows a clear excess of t_i at heterozygous sites in our ancient sample (**Fig. S26**). To correct for this bias, we estimated the excess of transition at heterozygous sites from the data. We first computed the mean and the Standard Deviation (SD) of this metric for all our data excluding the ancient Irish dog. We subtracted the expected excess of heterozygous transitions (the mean+1SD computed on modern samples) in the Newgrange sample to its total number of heterozygous sites. We then computed heterozygosity as the ratio of the number of heterozygous sites / total number of sites called in the genome. **Figure S27** shows that the corrected and non-correct value for the Newgrange dogs are only slightly different demonstrating that damage play a small role in this analysis. Our results show that the heterozygosity in the Newgrange dog is much closer to village dogs or to Asian breeds than it is to modern breeds demonstrating that artificial selection in this sample was not as strong.

Runs of Homozygosity

We also computed Runs of homozygosity (RoH). This metric is a good proxy to study recent inbreeding as it allows for the detection of long portion of the genome that identical and thus most likely inherited by closely related individuals and does not depend on the average heterozygosity (which can be influenced by long term population size reduction rather than recent inbreeding). To do so we used the RoH detection pipeline implemented in plink following the guideline from (*116*).

We first pruned our data for LD using --indep 50 5 2 and removed low frequency markers using -–maf 0.05 and a minimum number of SNPs of 50 (--homozyg-density 50). We then used different value for the maximum number of heterozygous in a RoH (1, 2 and 3) to assess the effect of different level of heterozygosity within RoHs (*e.g.* induced by damage). The results of these analyses are presented in **Figure S28**.

Firstly, we can see that the only sample affected by the number of heterozygosity allowed in a RoH is the ancient sample. Nevertheless, this pattern has little effect on our results. Indeed, in all cases the modern and Asian breeds show a clear excess of long RoHs (>=3Mb) demonstrating that inbreeding in these population is very high. In comparison level of inbreeding in the ancient Irish dog are closer to free breeding village dog. This demonstrates that artificial selection in this Neolithic dog had very less impact

on its genome than what is found in modern breeds. In addition, this suggests that these dogs may not have been actively bred but was rather free to choose their mates, as mostly do modern village dogs.

Phenotypic information

We used a few well-known markers that have good genotypic – phenotype association. We first assessed the presence of markers associated with starch digestion (*117*). We found 3 out of 3 markers (chr16-10117660:G, chr16- 10135196:C and chr16-10143343:T) in the *MGAM* display congruent genotypes between modern dogs and the Newgrange dog. These were found at very high frequency in the modern dog genomes (chr16-10117660:G, freq=1; chr16- 10135196:C, freq=0.95; chr16-10143343:T, freq=1). However, the newgrange dog had the ancestral allele (same as wolves) at the only marker associated with starch digestion in the *SGLT1* gene (chr26-27964111:A). Nevertheless this SNP was found at lower frequency in modern dog genomes (0.71). Together this suggests that this dog was able to process starch more efficiently than wolves.

We also assessed the copy number (CN) of the *AMY2B* gene as a previous study demonstrated that an increased expression of this gene is associated with higher CN and improved starch digestion (*117*). To do so we computed read depth (RD) in the first exon of the gene and compared it to the genome RD (GRD). We only used the first exon of this gene because this was the sequence that was validated using real time PCR in (*117*). The coordinates of the first exon was obtained from Ensembl v83. We also searched for this exon, in canfam3.1, using blat on the UCSC website to assess its CN in the reference. We found that the exon was repeated 5 times across the genome of the reference sequence (corresponding to 10 copies; **Table S6**). Our samples were aligned with bwa (*40*), which reports only 1 location at random if a read can align, equally well, at multiple positions in the reference genome. Thus, to compute the CN per sample we computed the RD at each position of this exon, summing over all copies in the reference, and divided by the mean genome wide RD. We show that this exon is clearly under copy number variation (CNV) in dogs with value ranging from 4-19 copies (raw CN ~3.69-18.98). However, we found that the CN among Wolves and the Newgrange dog was non-variable with only 2 copies (raw CN ~1.68-2.56; **Figure S29**).

We also checked for well-known markers associated with hair type in dogs (*118*). We found none of the derived allele at SNPs in gene *FGF5* and *KRT71* that are associated with modern dogs hair type. We also found that the Newgrange dog carried the ancestral allele at the *Fgf4* gene associated with chondrodysplasia (*119*).

Fig. S1.
Petrous bone of the Newgrange dog.

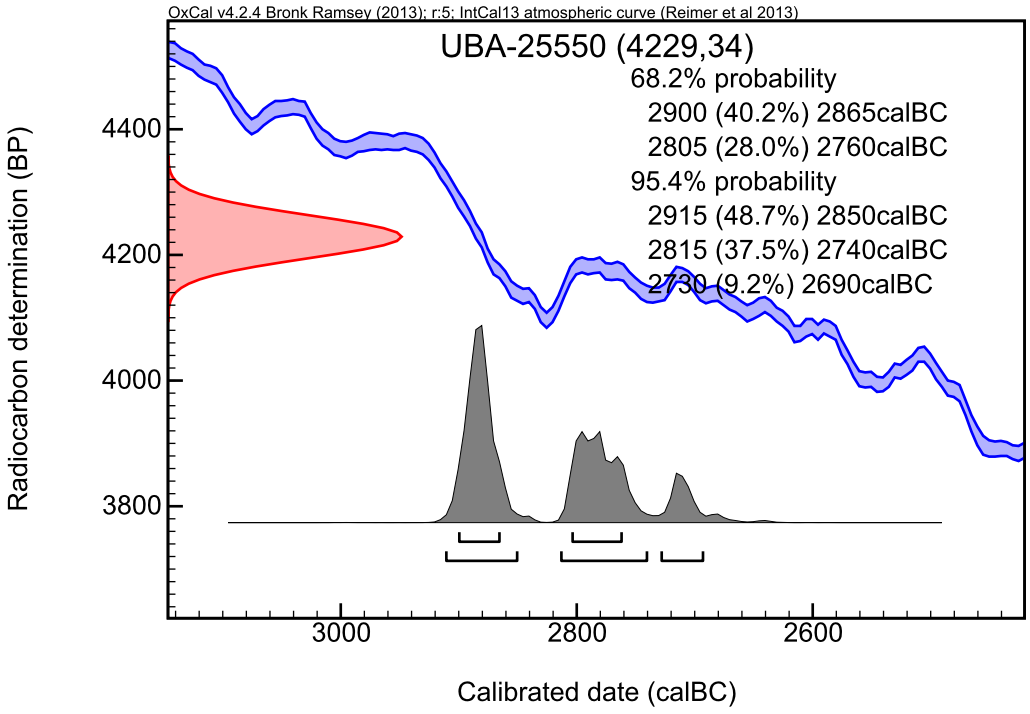


Fig. S2.

Calibration of the radiocarbon age of the newgrange dog.

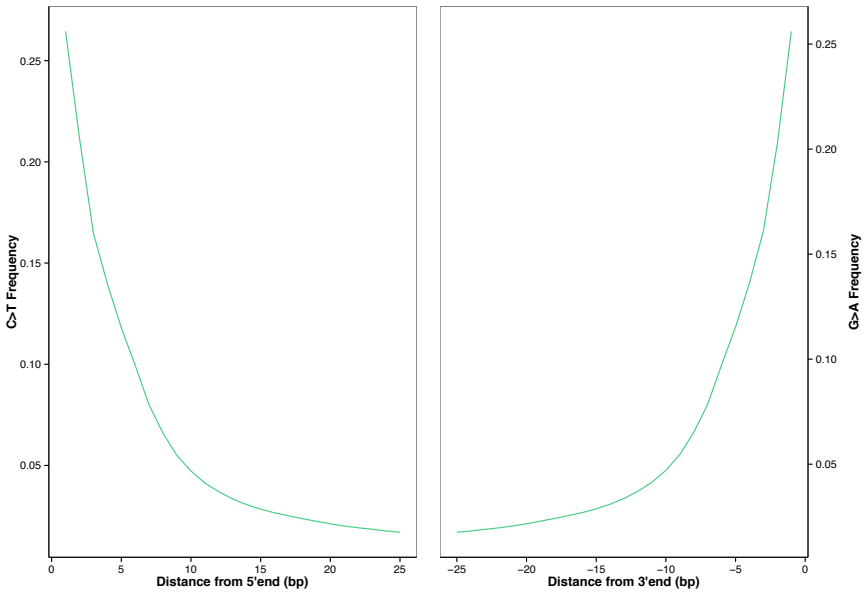


Fig. S3.
Deamination pattern in ancient reads.

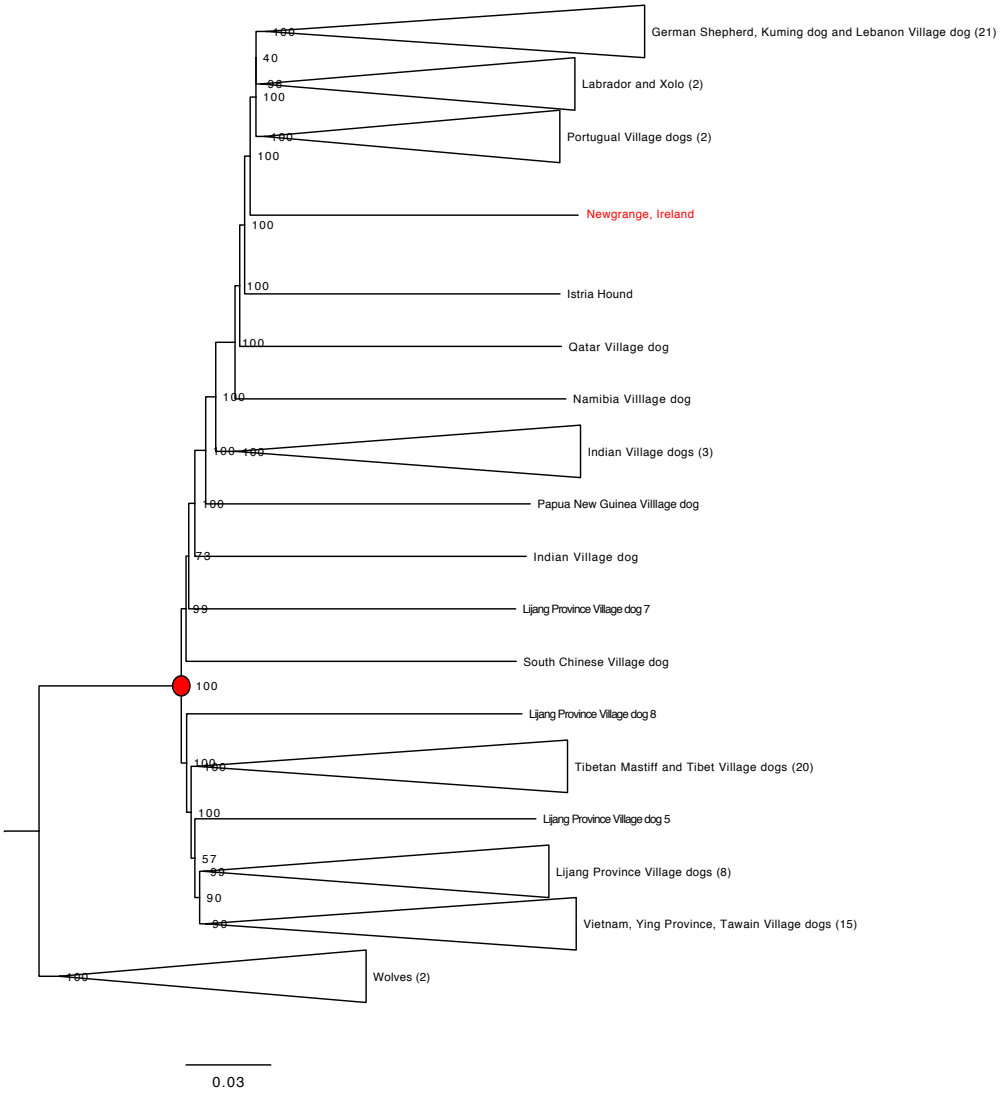


Fig. S4.

NJ tree based on IBS computed for 80 genome sequences of dogs and wolves. Red dot highlights the deep East Asian Western Eurasian split.

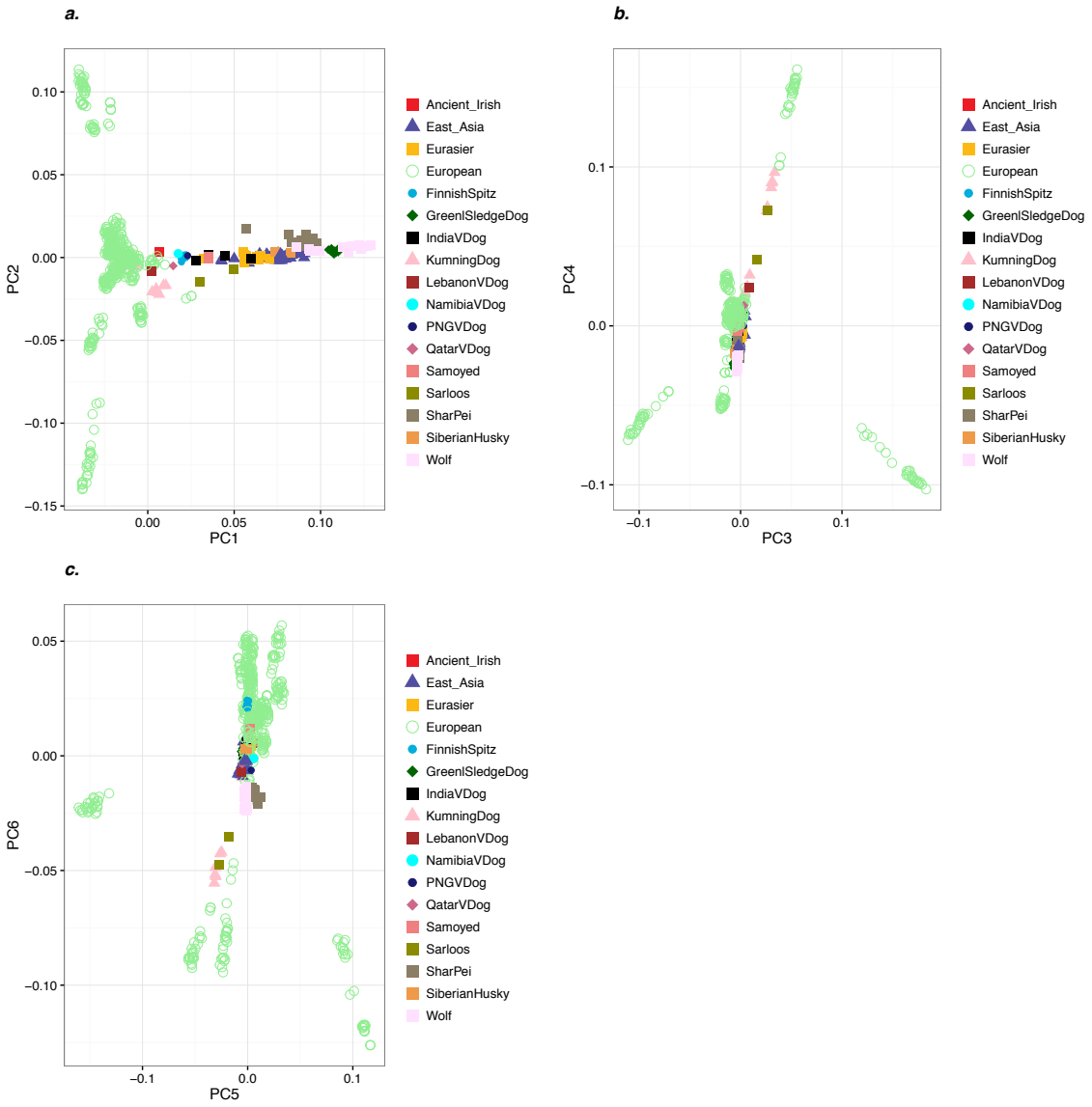
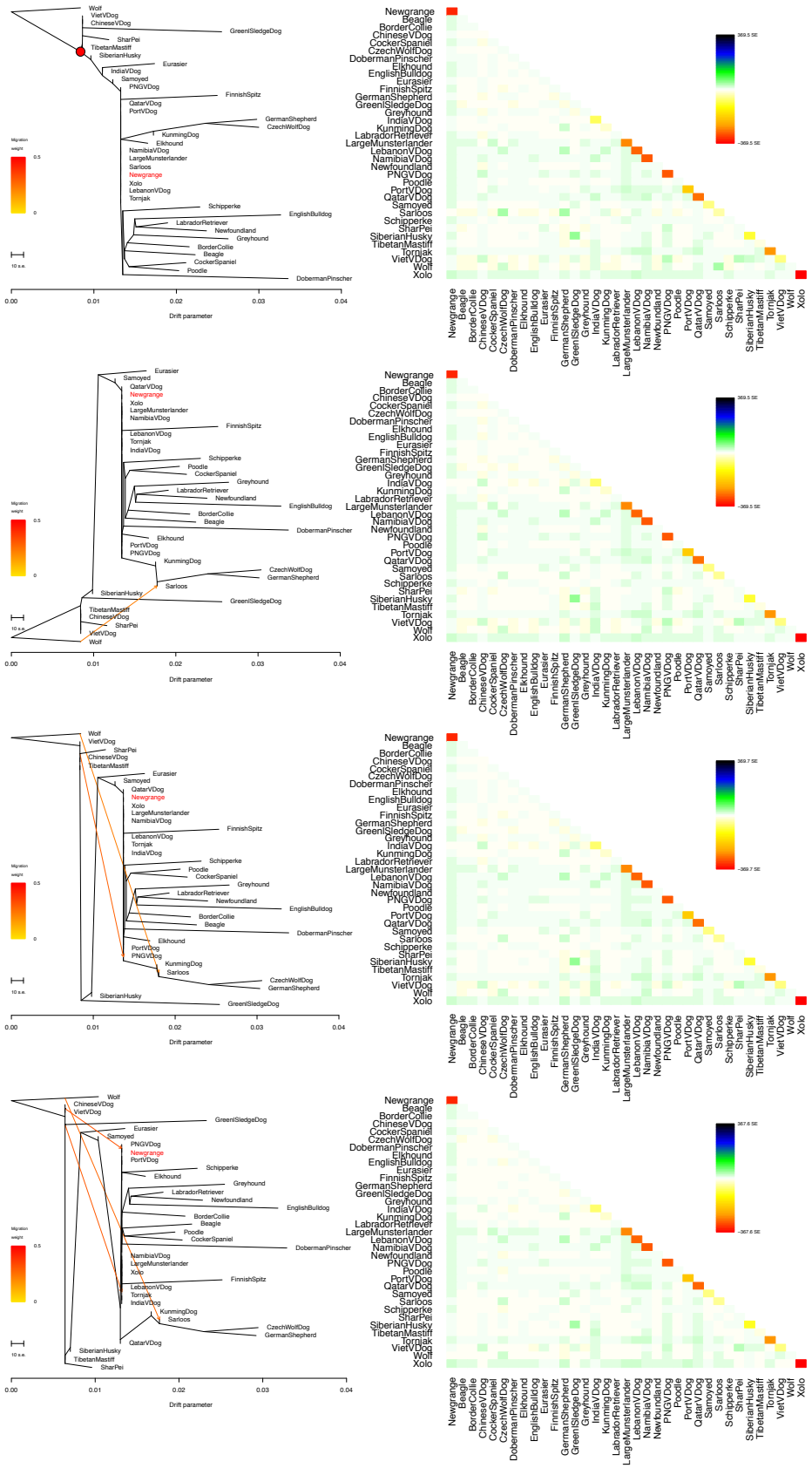


Fig. S5.

PCA based on 605 individuals.


Wolf
VietVDog
ChineseVDog
GreenlSledgeDog
SharPei
TibetanMastiff
SiberianHusky
Eurasier
IndiaVDog
Samoyed
PNGVDog
FinnishSpitz
QatarVDog
PortVDog
GermanShepherd
CzechWolfDog
KunmingDog
Elkhound
NamibiaVDog
LargeMunsterlander
Sarloos
Newgrange
Xolo
LebanonVDog
Tornjak
Schipperke
EnglishBulldog
LabradorRetriever
Newfoundland
Greyhound
BorderCollie
Beagle
CockerSpaniel
Poodle
DobermanPinscher
Migration weight
0.5
0
10 s.e.
0.00
0.01
0.02
0.03
0.04
Drift parameter
Newgrange
Beagle
BorderCollie
ChineseVDog
CockerSpaniel
CzechWolfDog
DobermanPinscher
Elkhound
EnglishBulldog
Eurasier
FinnishSpitz
GermanShepherd
GreenlSledgeDog
Greyhound
IndiaVDog
KunmingDog
LabradorRetriever
LargeMunsterlander
LebanonVDog
NamibiaVDog
Newfoundland
PNGVDog
Poodle
PortVDog
QatarVDog
Samoyed
Sarloos
Schipperke
SharPei
SiberianHusky
TibetanMastiff
Tornjak
VietVDog
Wolf
Xolo
369.5 SE
-369.5 SE

Eurasier
Samoyed
QatarVDog
Newgrange
Xolo
LargeMunsterlander
NamibiaVDog
FinnishSpitz
LebanonVDog
Tornjak
IndiaVDog
Schipperke
Poodle
CockerSpaniel
Greyhound
LabradorRetriever
Newfoundland
EnglishBulldog
BorderCollie
Beagle
DobermanPinscher
Elkhound
PortVDog
PNGVDog
KunmingDog
CzechWolfDog
GermanShepherd
Sarloos
SiberianHusky
GreenlSledgeDog
TibetanMastiff
ChineseVDog
SharPei
VietVDog
Wolf
Migration weight
0.5
0
10 s.e.
0.00
0.01
0.02
0.03
0.04
Drift parameter
Newgrange
Beagle
BorderCollie
ChineseVDog
CockerSpaniel
CzechWolfDog
DobermanPinscher
Elkhound
EnglishBulldog
Eurasier
FinnishSpitz
GermanShepherd
GreenlSledgeDog
Greyhound
IndiaVDog
KunmingDog
LabradorRetriever
LargeMunsterlander
LebanonVDog
NamibiaVDog
Newfoundland
PNGVDog
Poodle
PortVDog
QatarVDog
Samoyed
Sarloos
Schipperke
SharPei
SiberianHusky
TibetanMastiff
Tornjak
VietVDog
Wolf
Xolo
369.5 SE
-369.5 SE

Wolf
VietVDog
SharPei
ChineseVDog
TibetanMastiff
Eurasier
Samoyed
QatarVDog
Newgrange
Xolo
LargeMunsterlander
NamibiaVDog
FinnishSpitz
LebanonVDog
Tornjak
IndiaVDog
Schipperke
Poodle
CockerSpaniel
Greyhound
LabradorRetriever
Newfoundland
EnglishBulldog
BorderCollie
Beagle
DobermanPinscher
Elkhound
PortVDog
PNGVDog
KunmingDog
Sarloos
CzechWolfDog
GermanShepherd
SiberianHusky
GreenlSledgeDog
Migration weight
0.5
0
10 s.e.
0.00
0.01
0.02
0.03
0.04
Drift parameter
Newgrange
Beagle
BorderCollie
ChineseVDog
CockerSpaniel
CzechWolfDog
DobermanPinscher
Elkhound
EnglishBulldog
Eurasier
FinnishSpitz
GermanShepherd
GreenlSledgeDog
Greyhound
IndiaVDog
KunmingDog
LabradorRetriever
LargeMunsterlander
LebanonVDog
NamibiaVDog
Newfoundland
PNGVDog
Poodle
PortVDog
QatarVDog
Samoyed
Sarloos
Schipperke
SharPei
SiberianHusky
TibetanMastiff
Tornjak
VietVDog
Wolf
Xolo
369.7 SE
-369.7 SE

Wolf
ChineseVDog
VietVDog
GreenlSledgeDog
Eurasier
Samoyed
PNGVDog
Newgrange
PortVDog
Schipperke
Elkhound
Greyhound
LabradorRetriever
Newfoundland
EnglishBulldog
BorderCollie
Beagle
Poodle
CockerSpaniel
DobermanPinscher
NamibiaVDog
LargeMunsterlander
Xolo
FinnishSpitz
LebanonVDog
Tornjak
IndiaVDog
KunmingDog
Sarloos
CzechWolfDog
GermanShepherd
QatarVDog
SiberianHusky
TibetanMastiff
SharPei
Migration weight
0.5
0
10 s.e.
0.00
0.01
0.02
0.03
0.04
Drift parameter
Newgrange
Beagle
BorderCollie
ChineseVDog
CockerSpaniel
CzechWolfDog
DobermanPinscher
Elkhound
EnglishBulldog
Eurasier
FinnishSpitz
GermanShepherd
GreenlSledgeDog
Greyhound
IndiaVDog
KunmingDog
LabradorRetriever
LargeMunsterlander
LebanonVDog
NamibiaVDog
Newfoundland
PNGVDog
Poodle
PortVDog
QatarVDog
Samoyed
Sarloos
Schipperke
SharPei
SiberianHusky
TibetanMastiff
Tornjak
VietVDog
Wolf
Xolo
367.6 SE
-367.6 SE

Fig. S6.

Results of *TreeMix* analyses with and without migration edge, red square represents the East Asian / Western Eurasian in Fig. 1a.

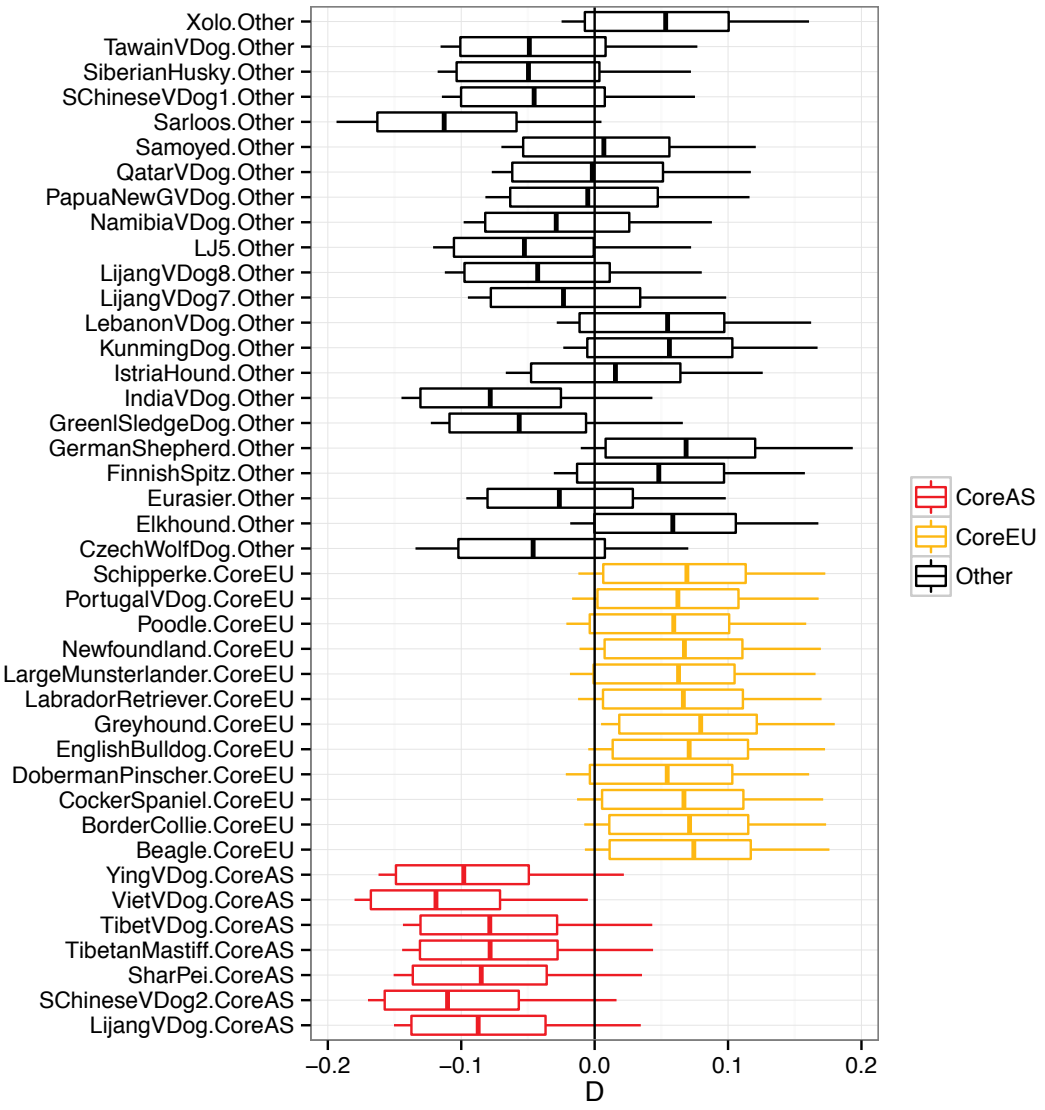


Fig. S7.

D-statistics support the affinity of the Newgrange dog with Modern European dogs. Here we computed D(Outgroup, aIrish; A, B), where A and B represent every possible pair of populations. The Y axis represent population A. Positive values support a close relationship between A and the Newgrange dog while negative support a close relationship between B and the Newgrange dog.

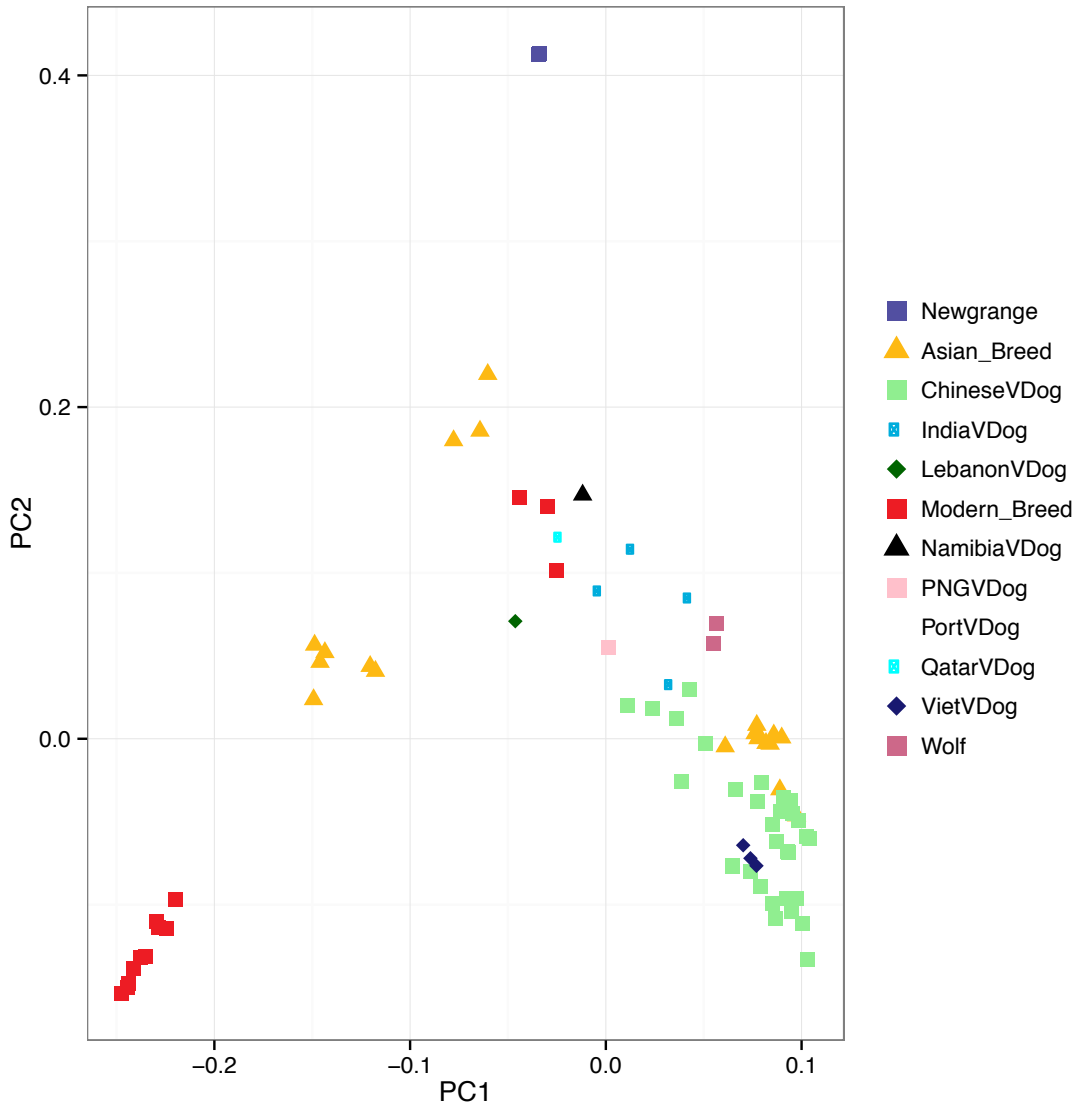


Fig. S8.

Principal component analysis based on 348,927 SNPs ascertained solely in the genome-wide data-set. While it is impossible to distinguish there are two purple squares for the Newgrange dog, with and without quality score re-calibration (see Material and Methods).

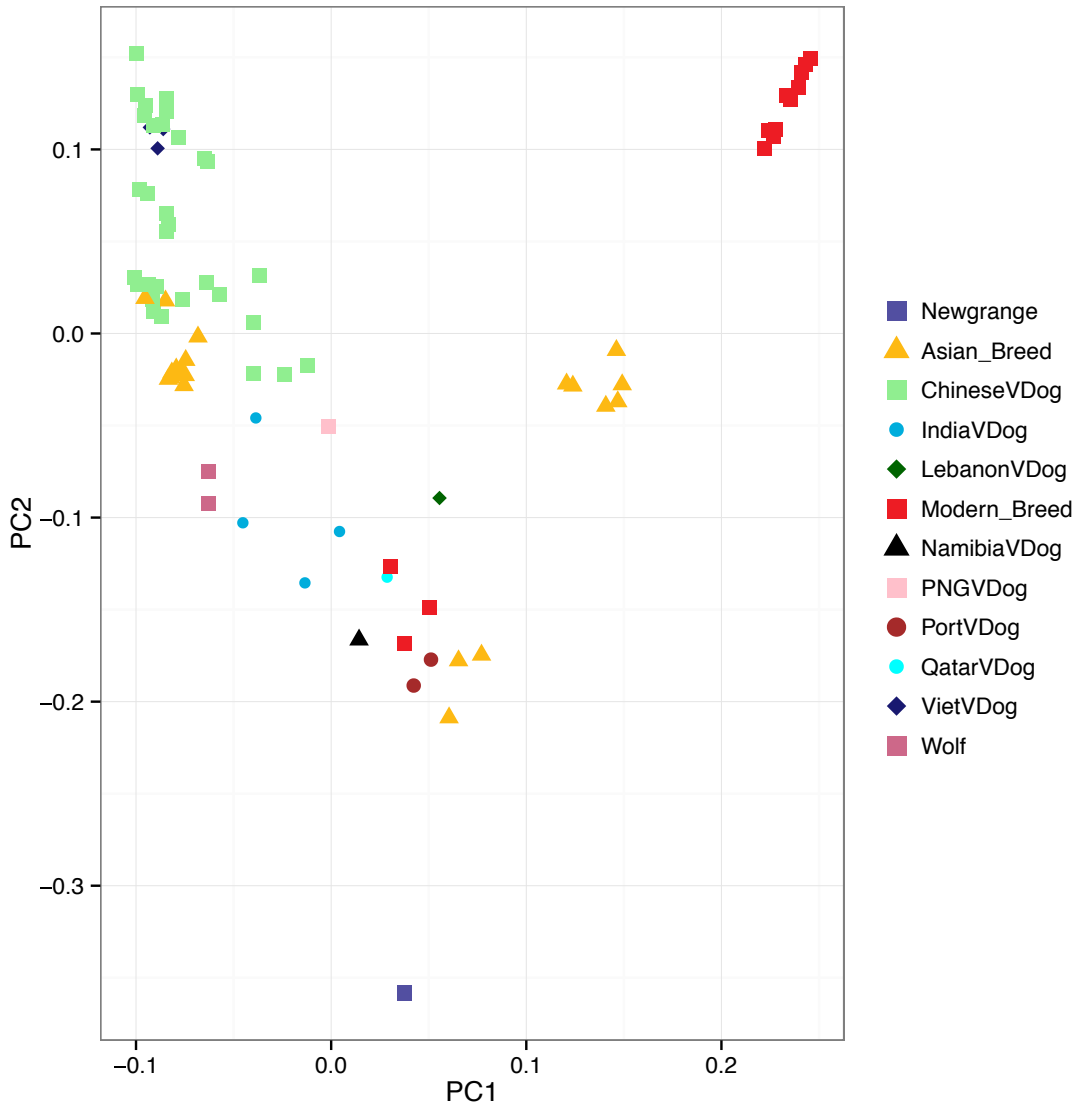


Fig. S9.

Principal component analysis based on 269,512 transversion ascertained solely in the genome-wide data-set. While it is impossible to distinguish there are two purple squares for the Newgrange dog, with and without quality score re-calibration.

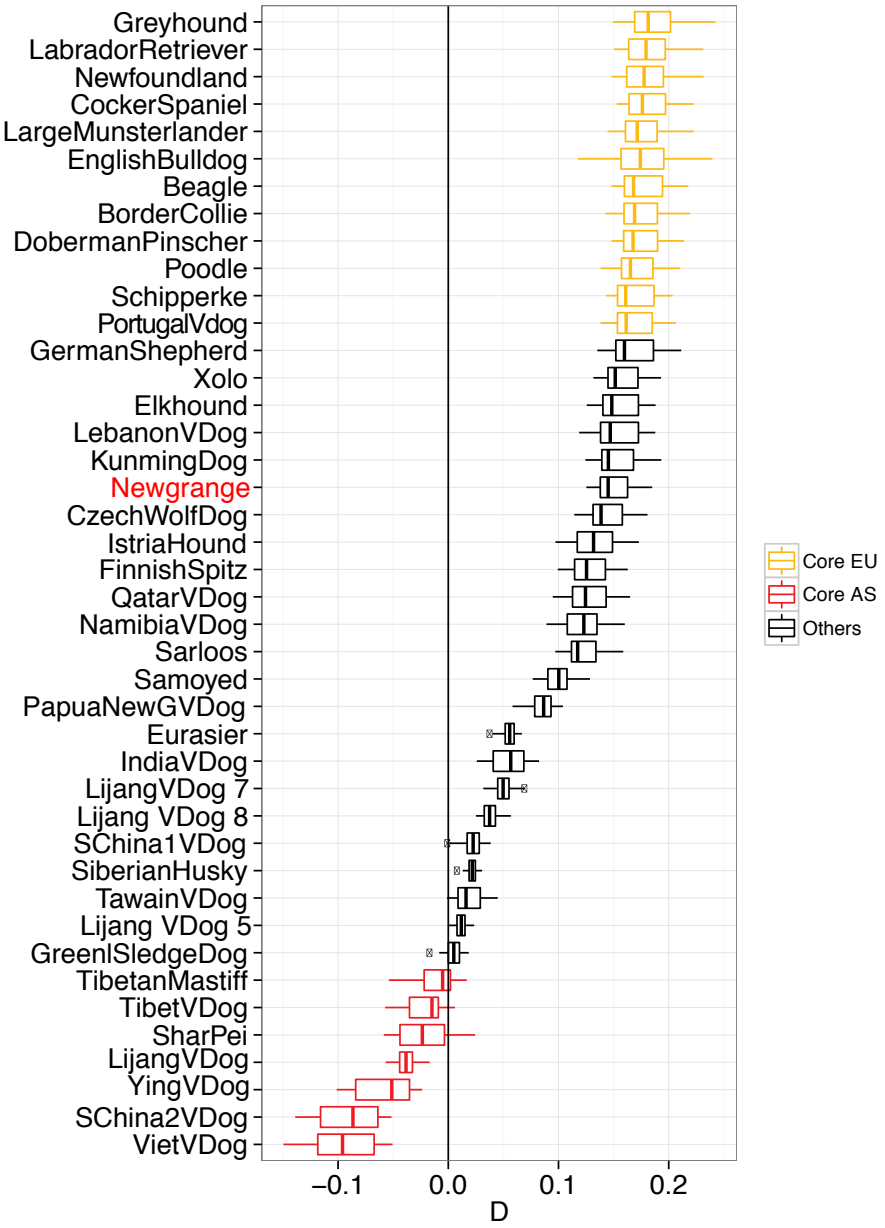


Fig. S10.

D-statistic tests representing the affinity of each dog population (or breed) to either Western Eurasian or East Asian core group. Negative values imply that the population on the Y-axis is closer to the East Asian core, positive values imply that the population is closer to the Western Eurasian core.

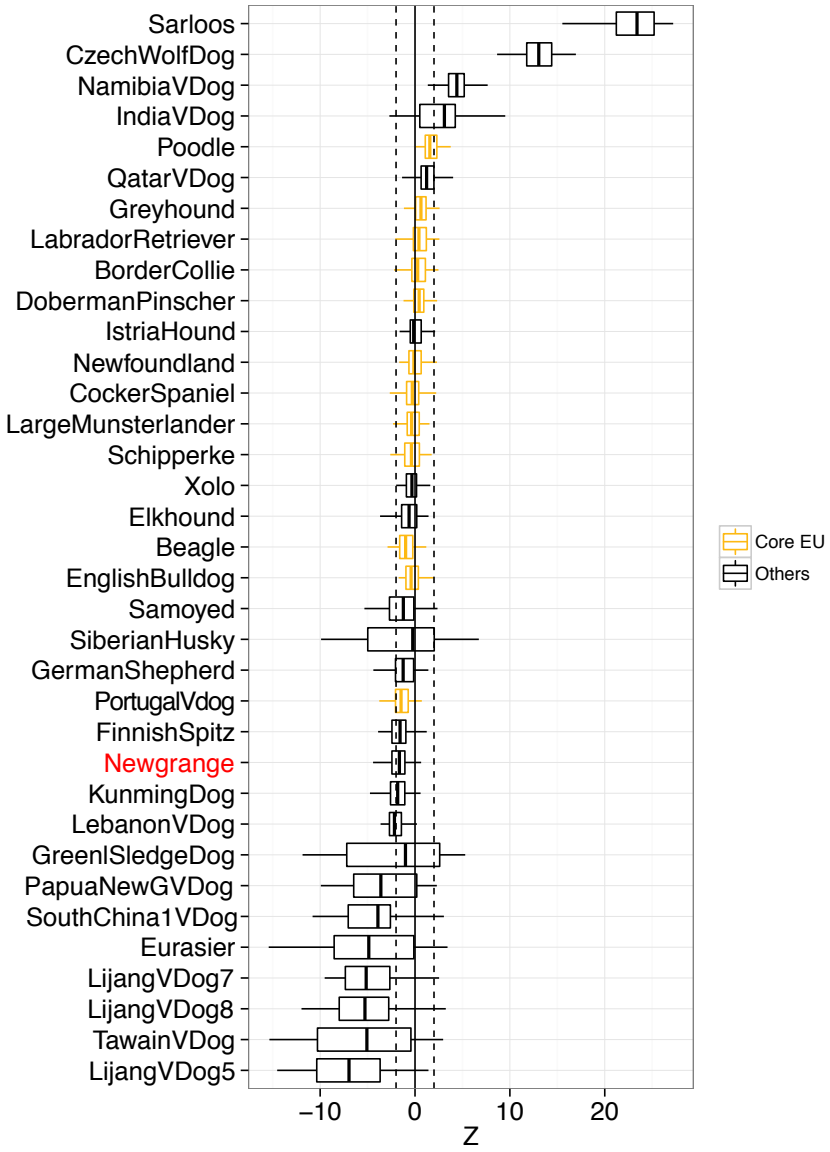


Fig. S11.

Significance of D tests (Z) representing mixed ancestry of a population (Y-axis). If a population is not been admixed with East Asian, we expect -2<Z<2 (X-axis). Populations admixed with East Asian have negative values, populations less admixed than others (*e.g.* due to admixture with wolves) will have positive values.

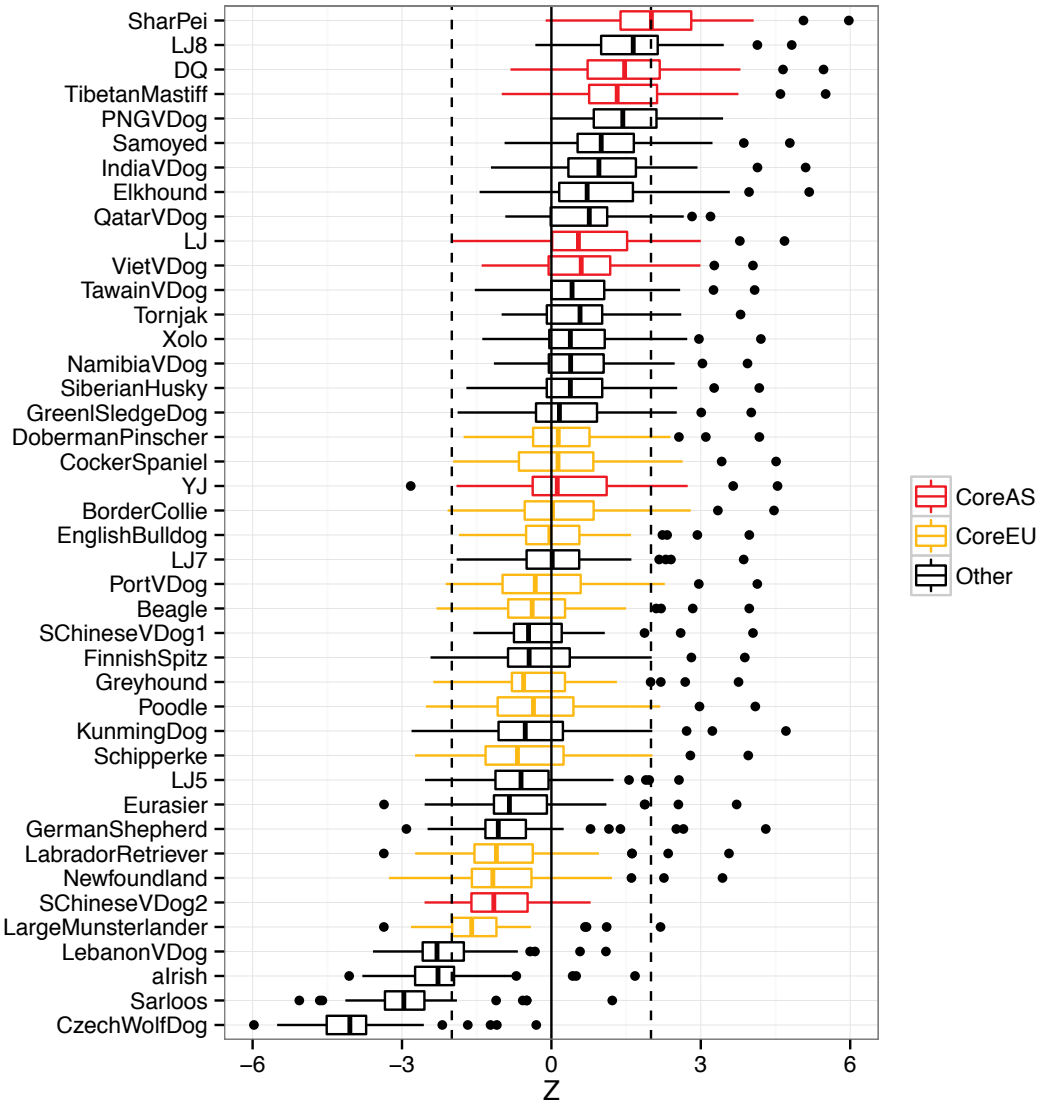


Fig. S12.

Significance of D tests (Z) representing admixture between wolves and various populations.

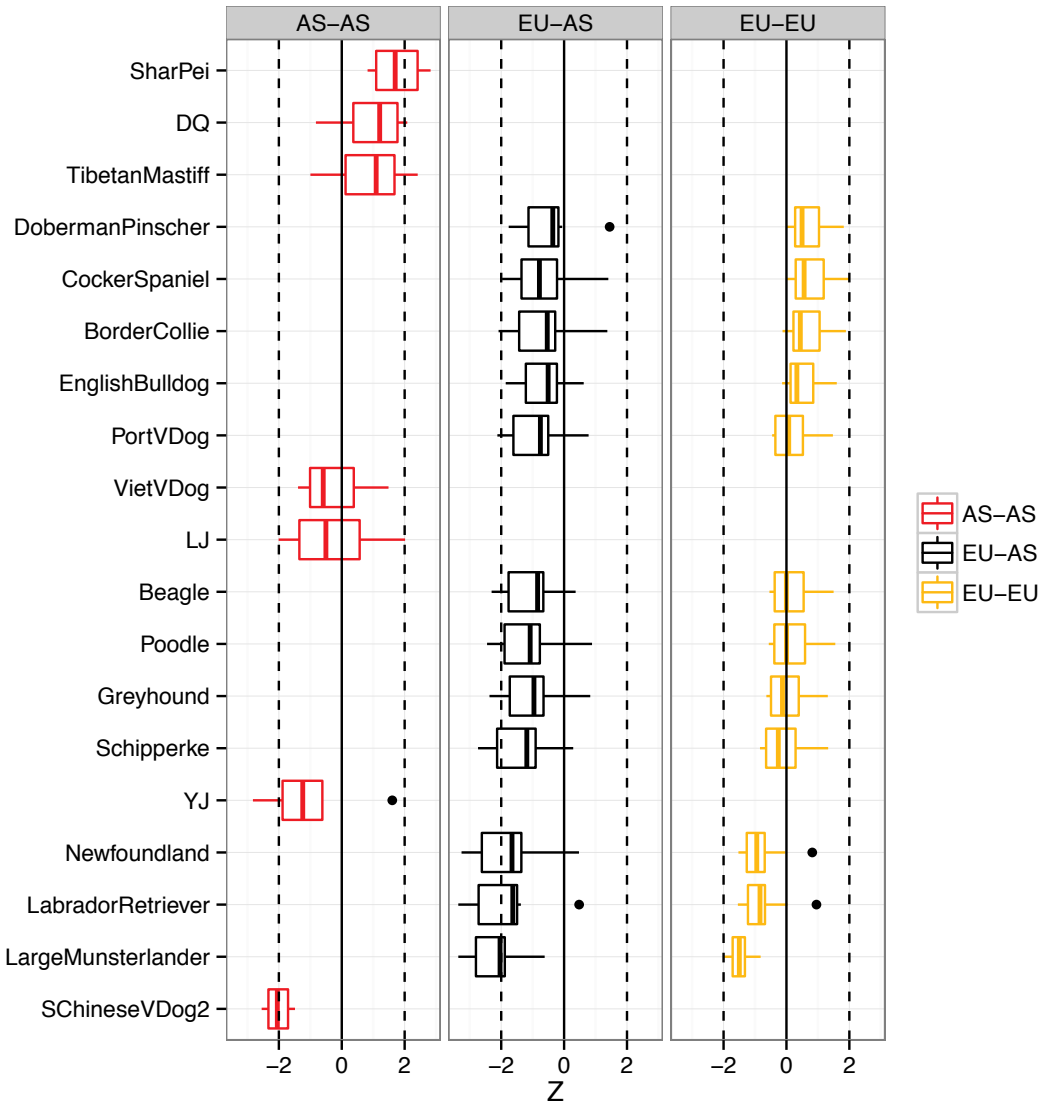


Fig. S13.

Significance of D tests (Z) representing admixture between wolves and core populations. The leftmost boxplots represent D(Outgroup, Wolves; East Asian, East Asian), the second D(Outgroup, Wolves; Western Eurasian, East Asian) or D(Outgroup, Wolves; East Asia, Western Eurasia) and the right plot represents D(Outgroup, Wolves; Western Eurasia, Western Eurasia).

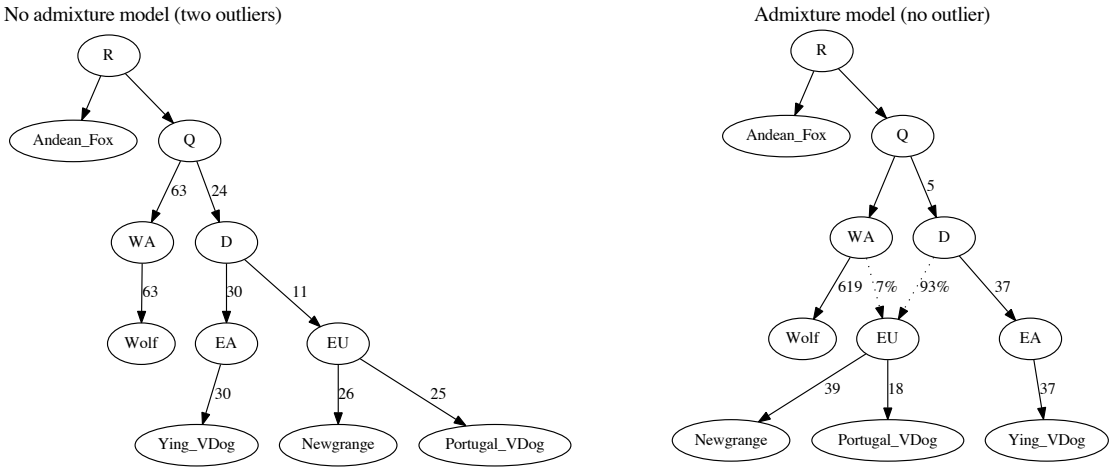


Fig. S14.

Two possible admixture graphs were fitted to the 170K SNP data, with and without admixture from wolves. The first graph on the left side left two f_4 outliers, while the second graph (right side) left none.

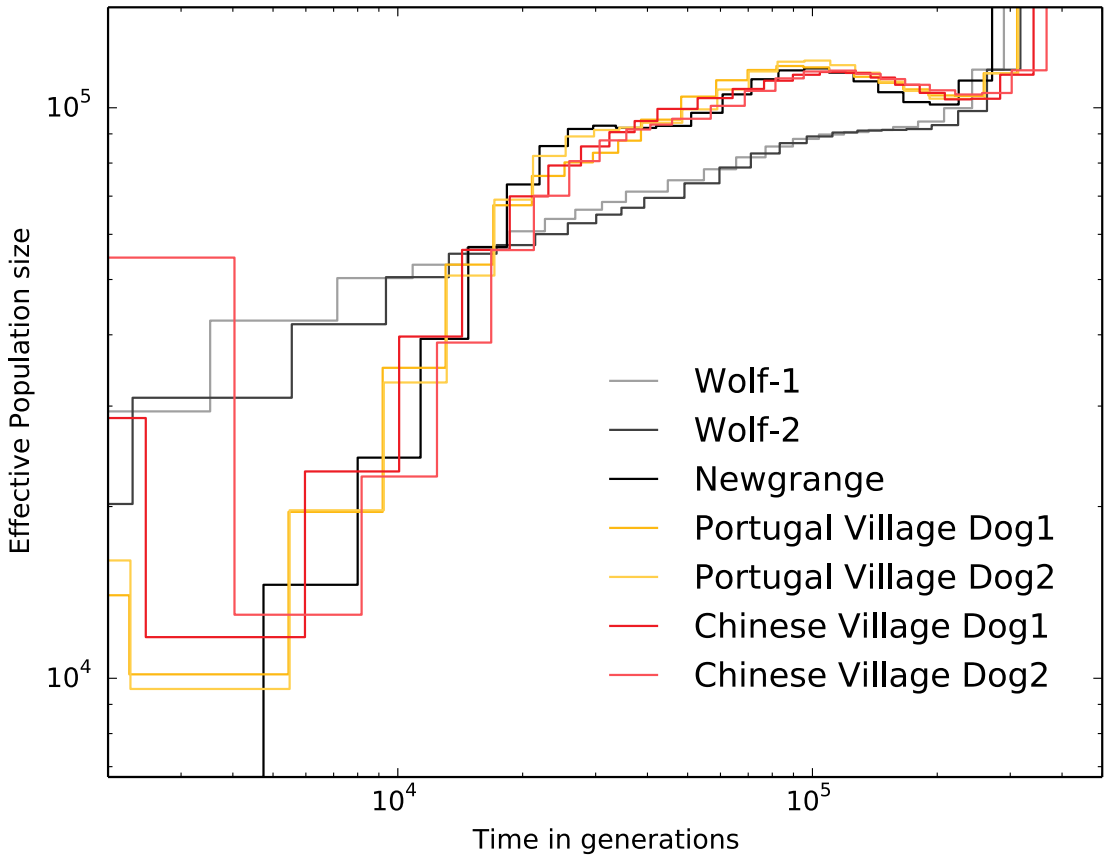


Fig. S15.

Effective population size through time of wolves, Newgrange dog and Portuguese/Chinese village dogs using PSMC' (single sample MSMC).

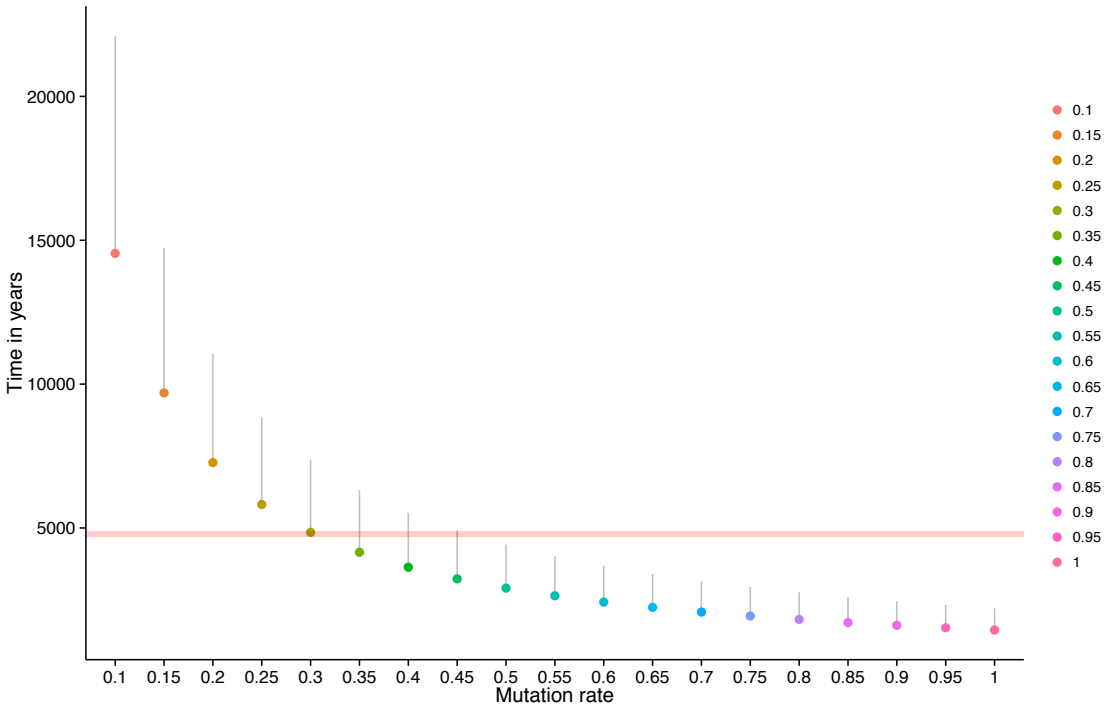


Fig. S16.

Canid mutation rate inferred using MSMC**.** Portuguese – Newgrange dog divergence time inferred by MSMC using various mutation rates. Dots, lower and upper bar represents the time at which cross-coalescence rate dropped below, 50%, 25% and 75% respectively. Red shaded area represents the radiocarbon age of the Newgrange dog.

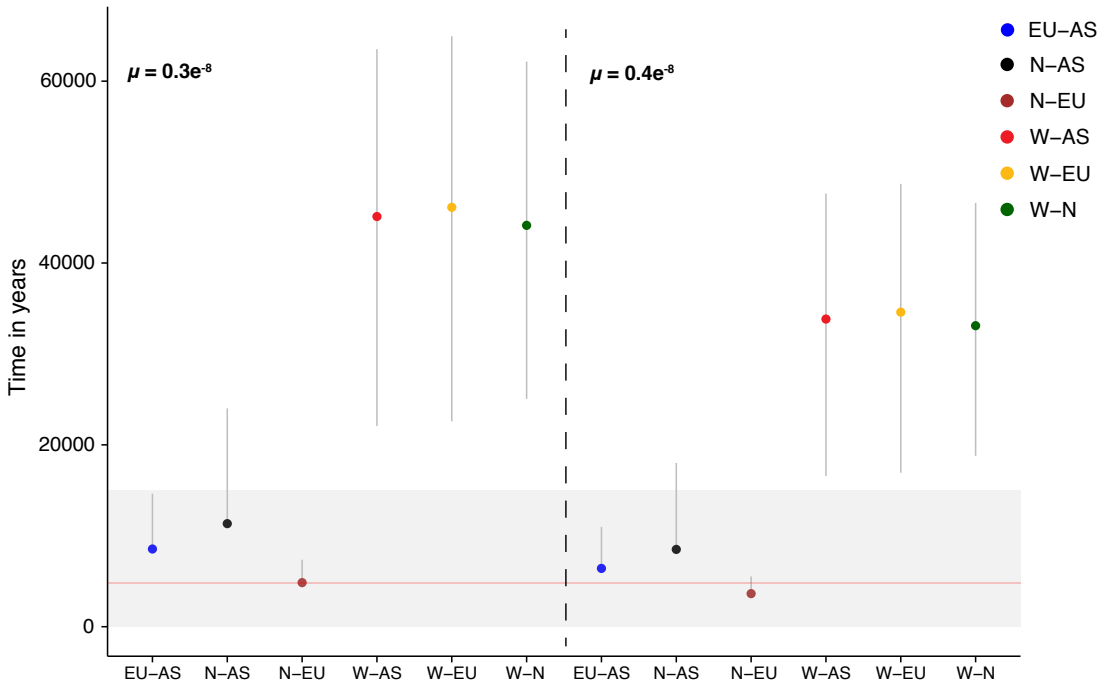


Fig. S17.

Divergence time between populations as inferred from MSMC with two different mutation rates calibrated using the Newgrange dog radiocarbon age. Dots, lower and upper bar represents the time at which cross-coalescence rate dropped below, 50%, 25% and 75% respectively. N=Newgrange, EU=Western Eurasian, AS=Eastern Asia, W=Wolf. The red bar represents the radiocarbon age of the Newgrange dog. The grey shaded area represents the time at which domestic dogs were present in Europe (~15,000 BP).

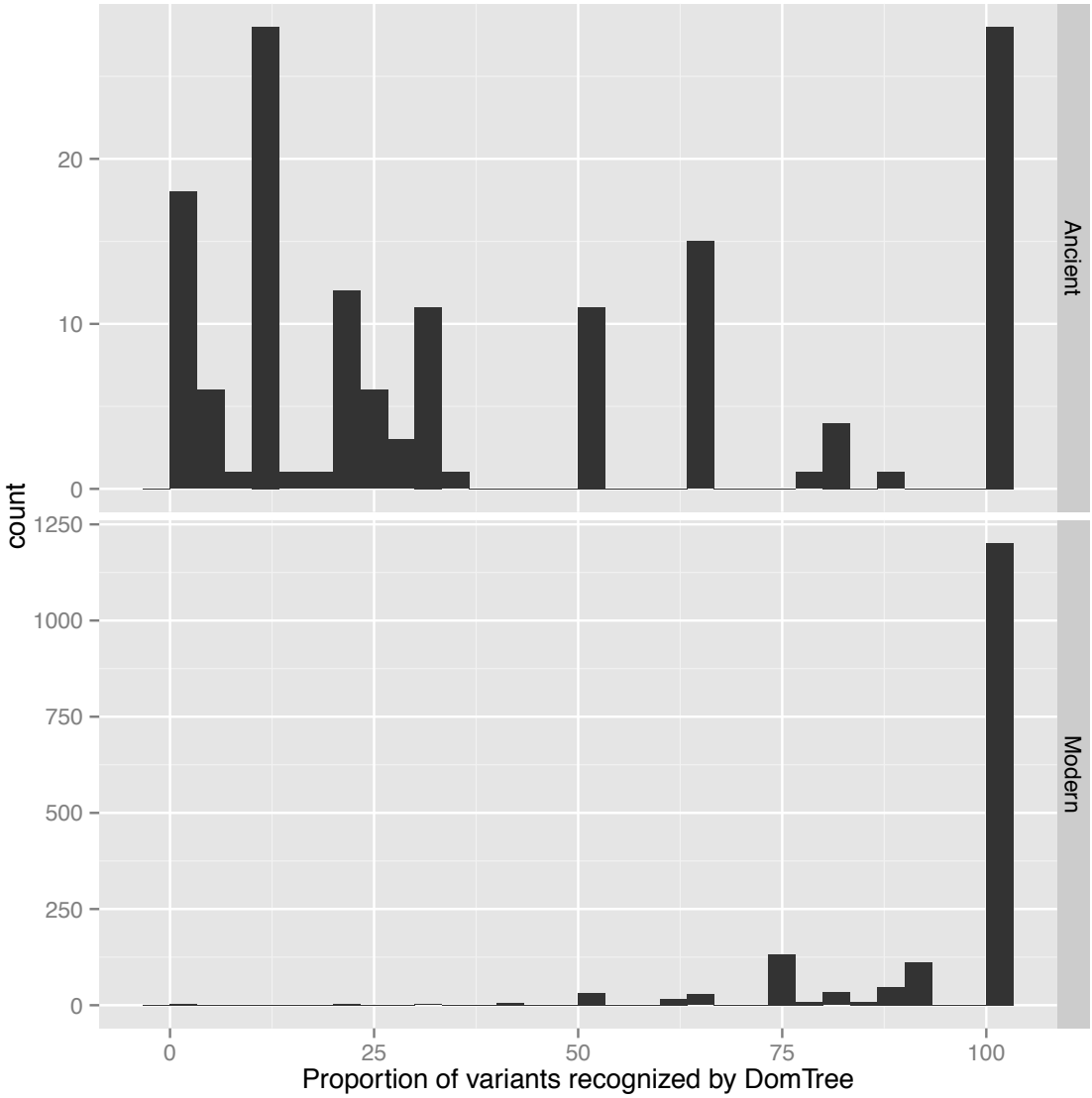


Fig. S18.

This histogram represents the proportion of variants, scored against reference genome, that were found in each of our sequence and were recognised by DomeTree's.

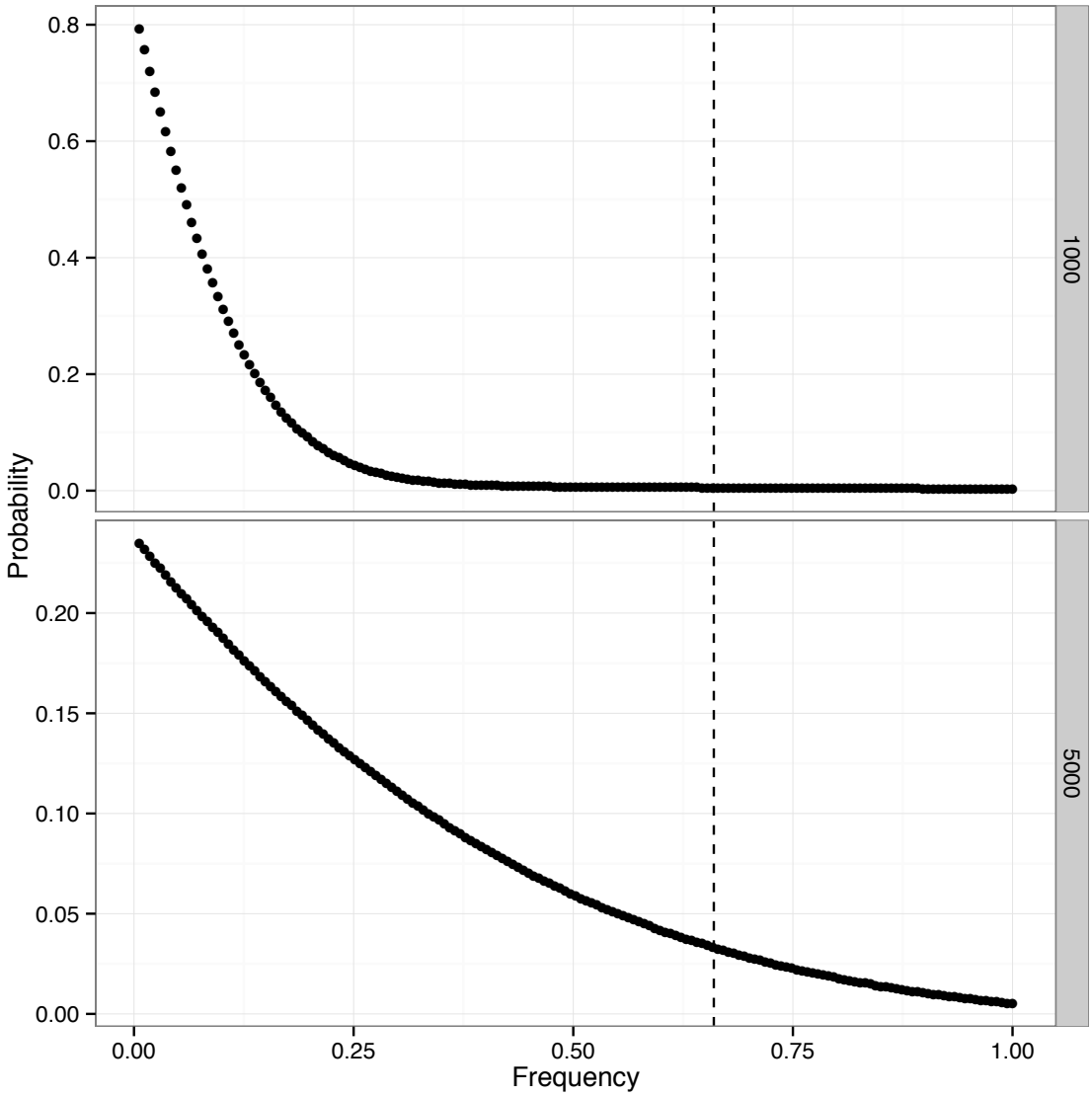


Fig. S19.

Probability of observing any given allele frequency (or higher) under our MSMC demographic model and only drift with a starting frequency equal to 0.1 (see text). Dashed line represents our estimate of the frequency of haplogroup A at t=0 (0.66) in the mtDNA data-set.

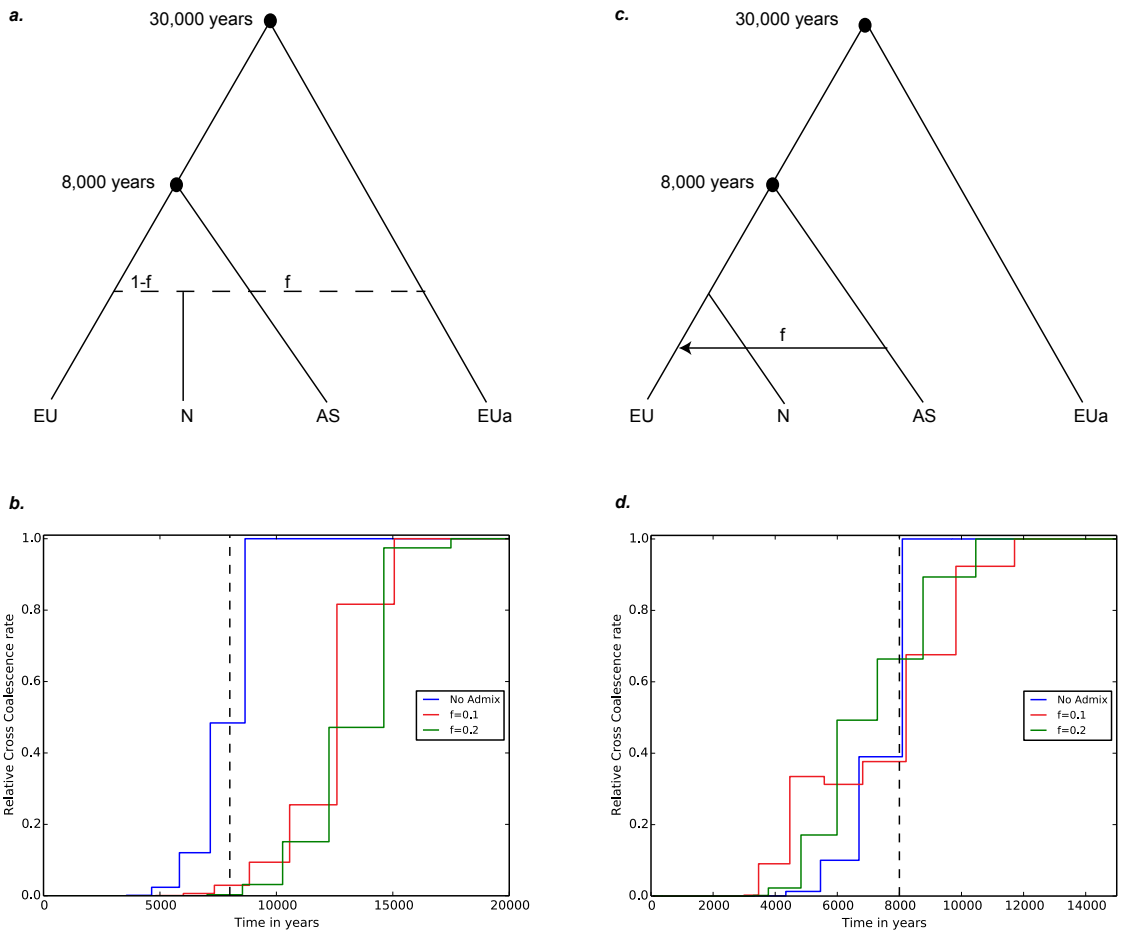


Fig. S20.

Diagrams and Cross-coalescence rates for simulated data under various scenarios of admixture. **a.** Partial replacement model **b.** secondary gene flow model. **c**. CCR for the partial replacement, where the Newgrange dog (N) derives f=0 (no admix), f=0.1 (10%) or f=0.2 (20%) of its ancestry from an ancient population (EUa) **d.** CCR for the secondary gene flow model, where modern European dogs (EU) admixed with Asian dogs (AS) after they split from the Newgrange dog.

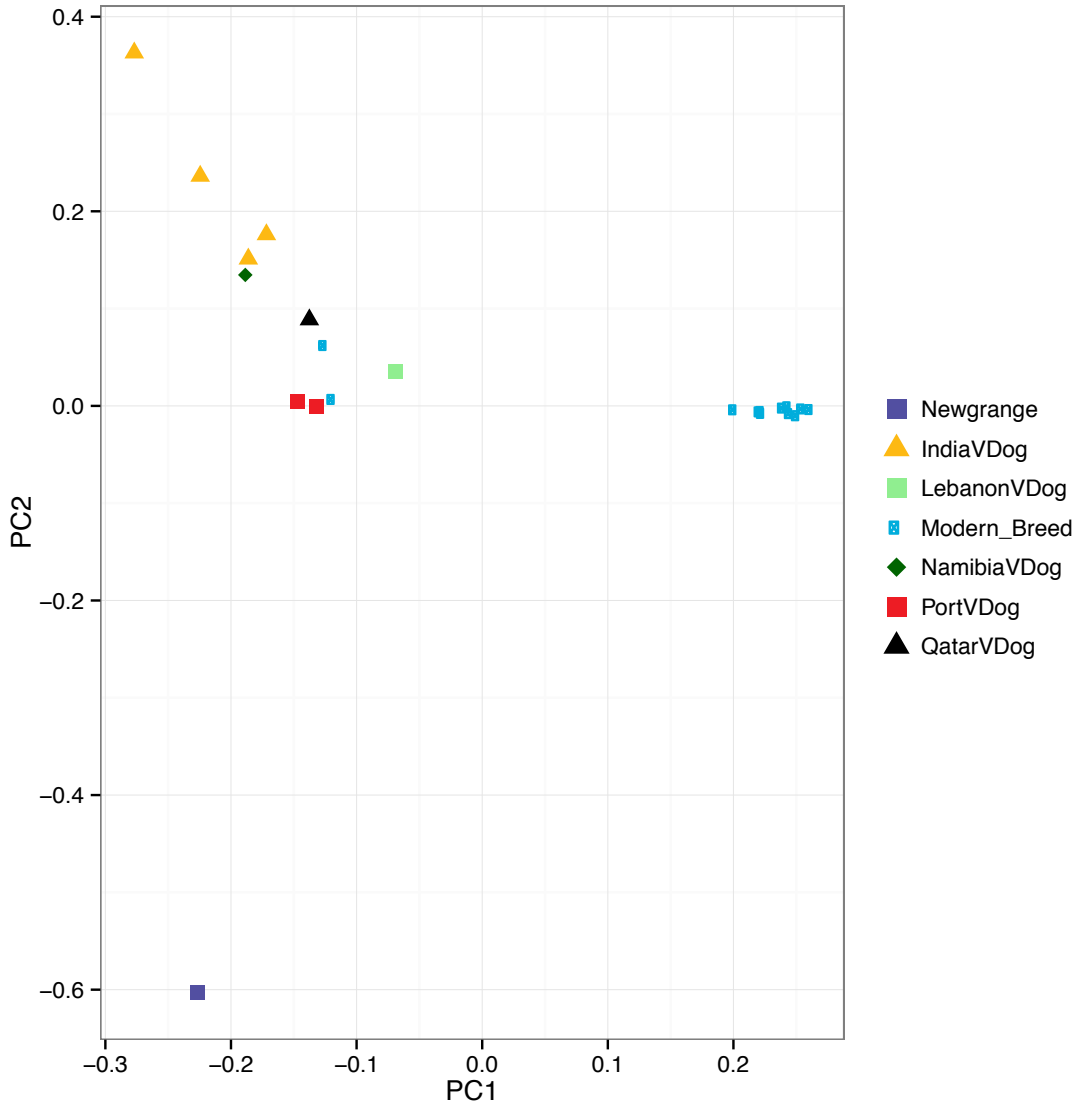


Fig. S21.

Principal component analysis of non-East Asian samples based on 242,745 transversions.

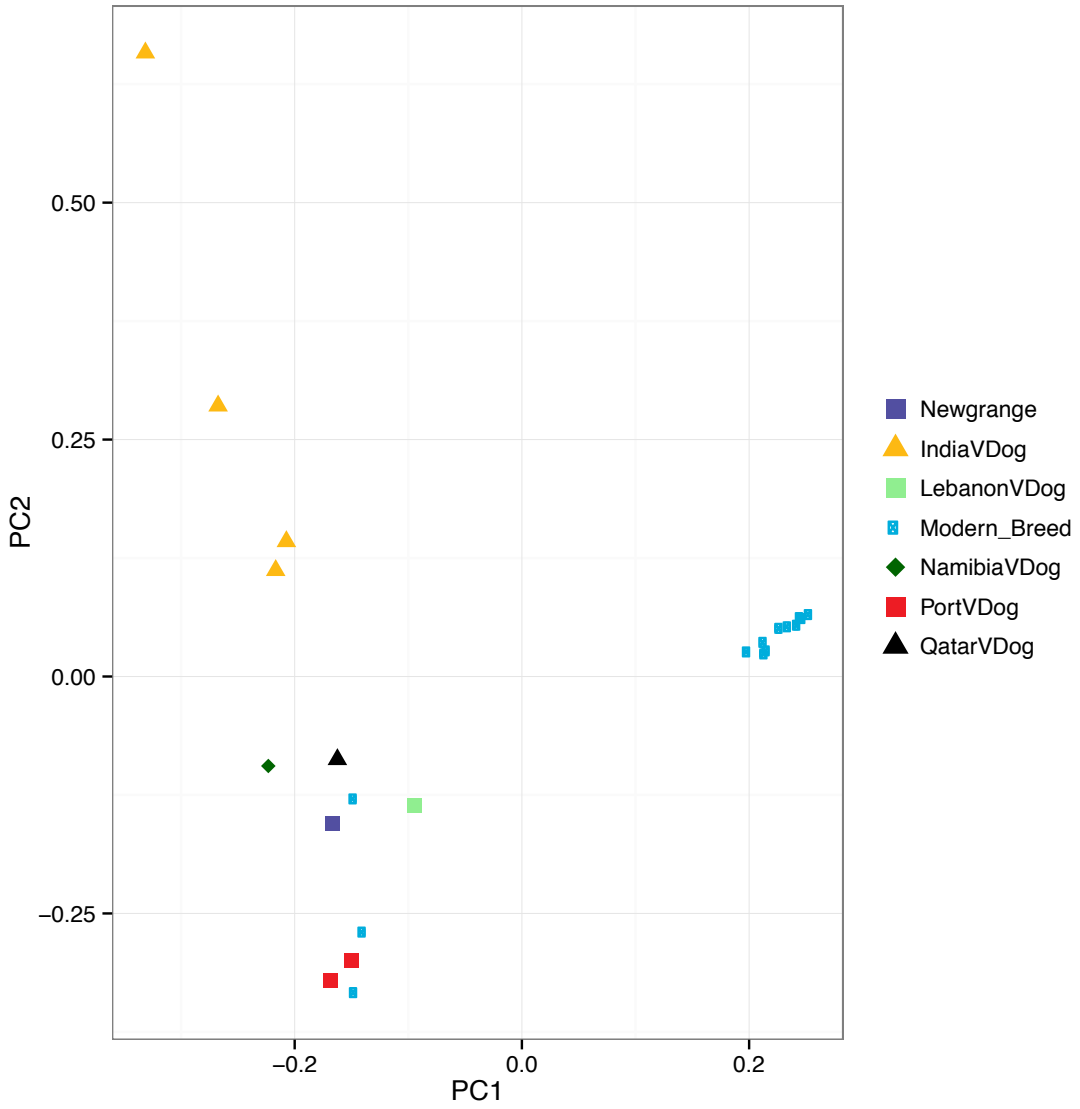


Fig. S22.

Principal component analysis of non-East Asian samples based on 242,745 transversions, without the Newgrange dog (projected).

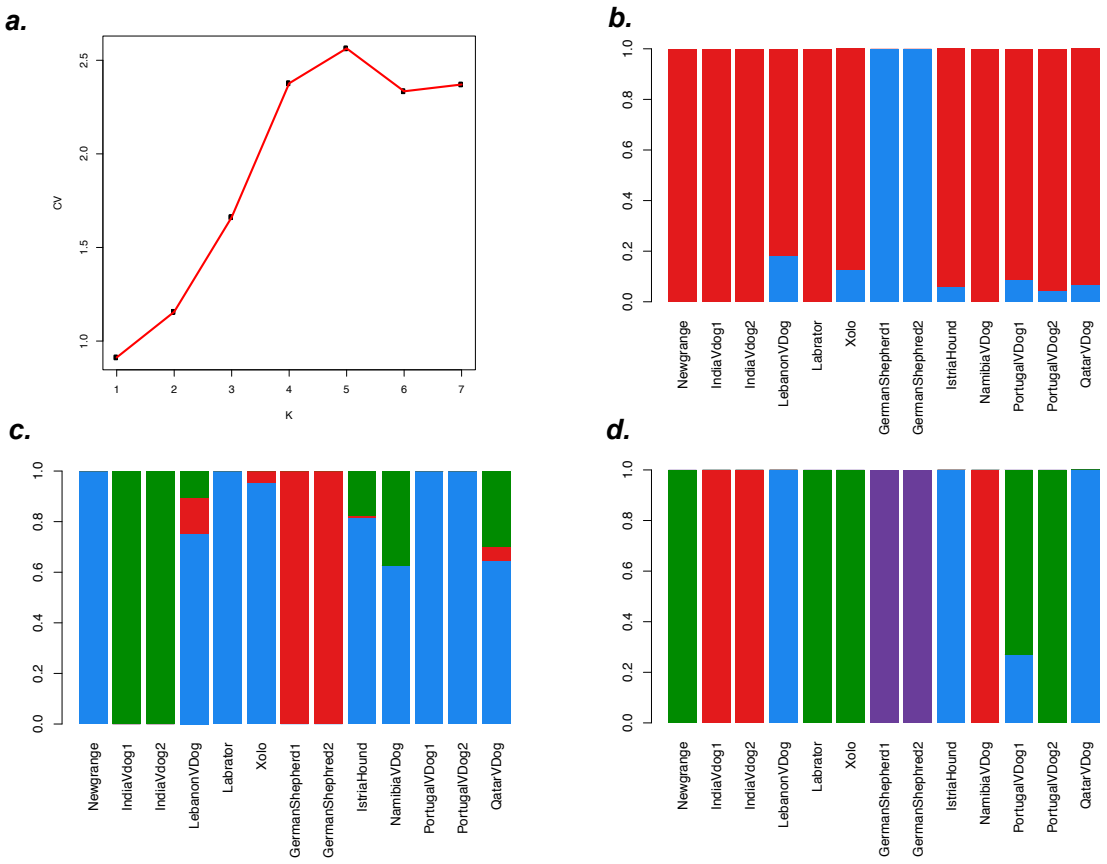


Fig. S23.

ADMIXTURE analysis of non-East Asian samples based on 242,745 transversions. **a.** Cross-validation values. **b.** K=2. **c.** K=3. **d.** K=4.

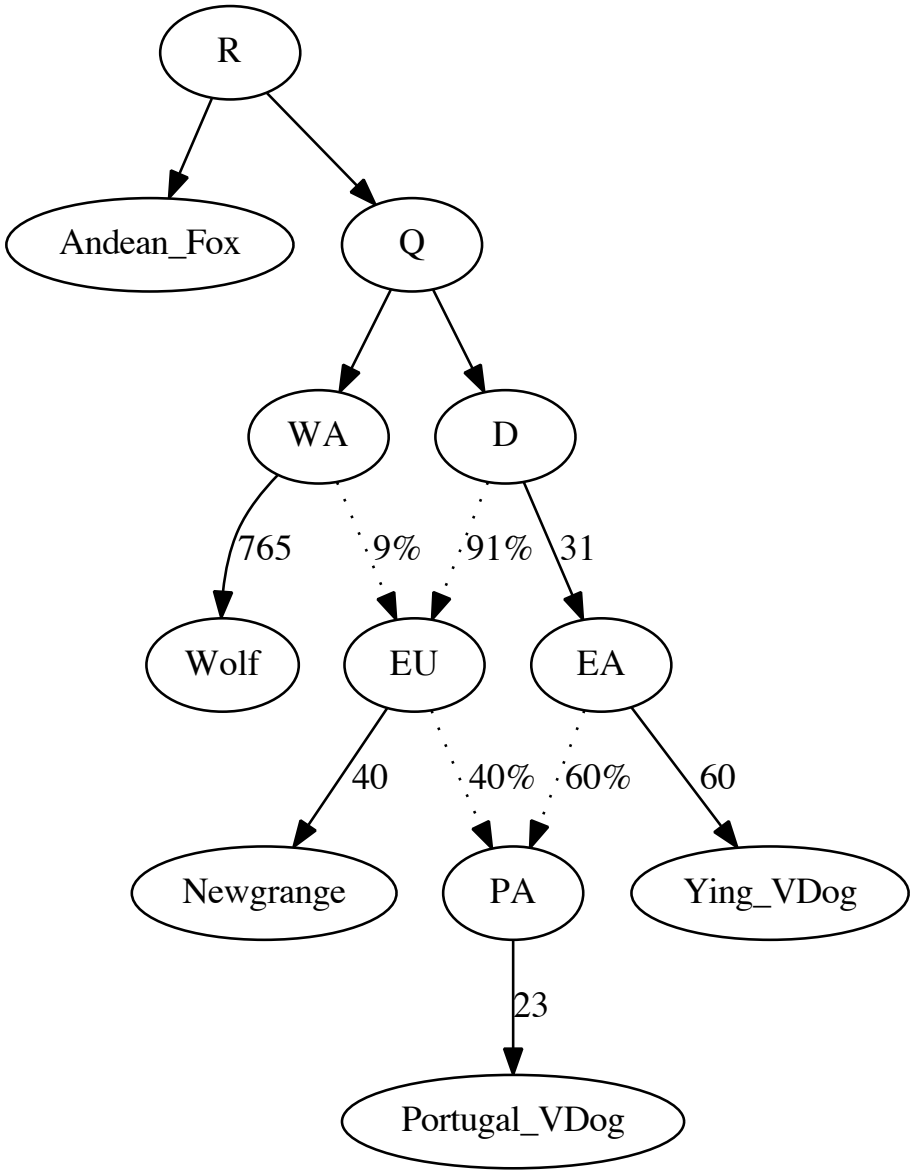


Fig. S24.
ADMIXTUREGRAPH model fitted with two wave of admixture from Asian dogs.

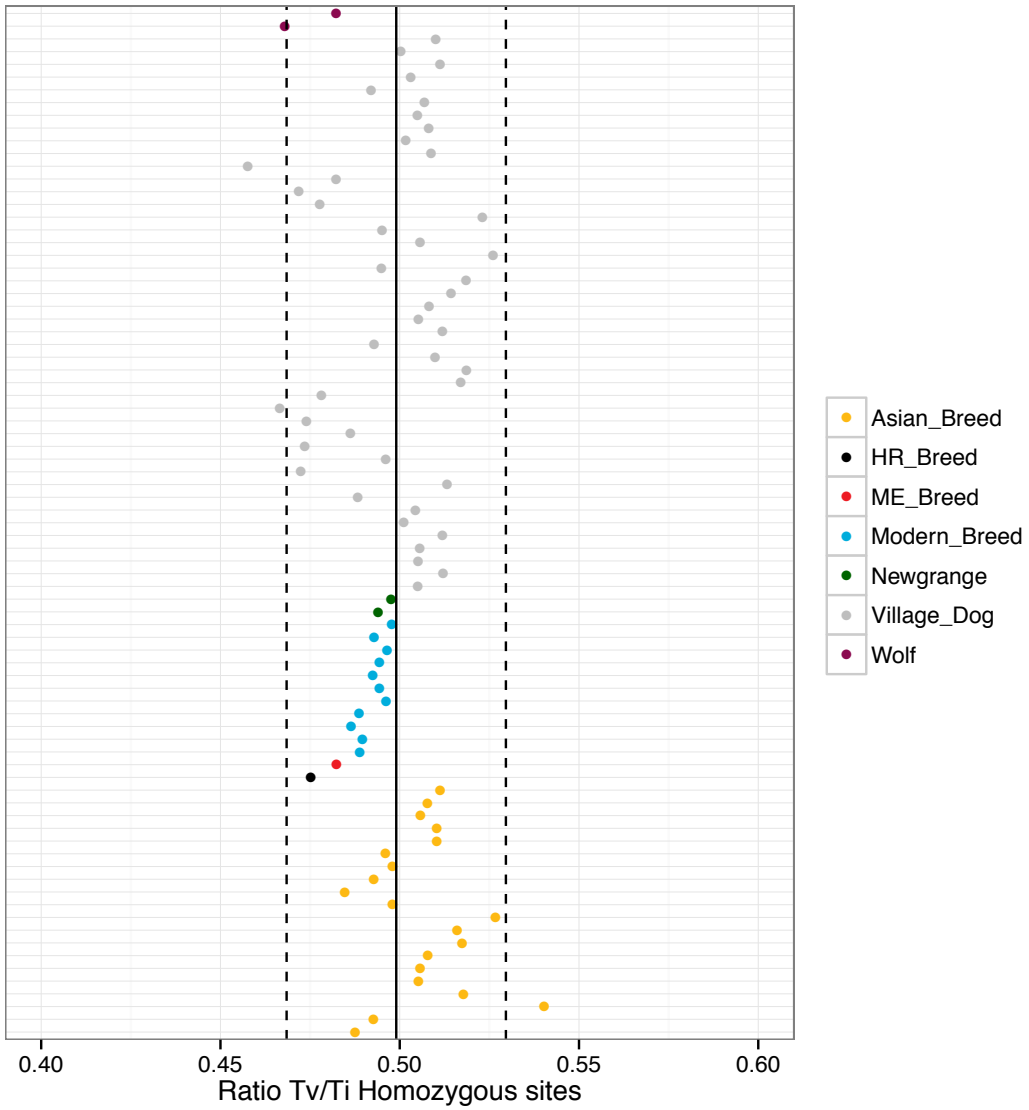


Fig. S25.

Proportion of Tv Ti at homozygous non-reference in each sample. The two green dots represents re-calibrated (with mapdamage; see above) and non re-calibrated respectively. Solid black line represent mean across all samples, dotted lines represent ±1SD. HR_breed=Istria dog; ME_breed=Xolo; Village dogs = all village dogs; Asian breeds= all Asian breeds (including Kunming dog); Modern breeds= German Shepherd, Tornjack and Labrador.

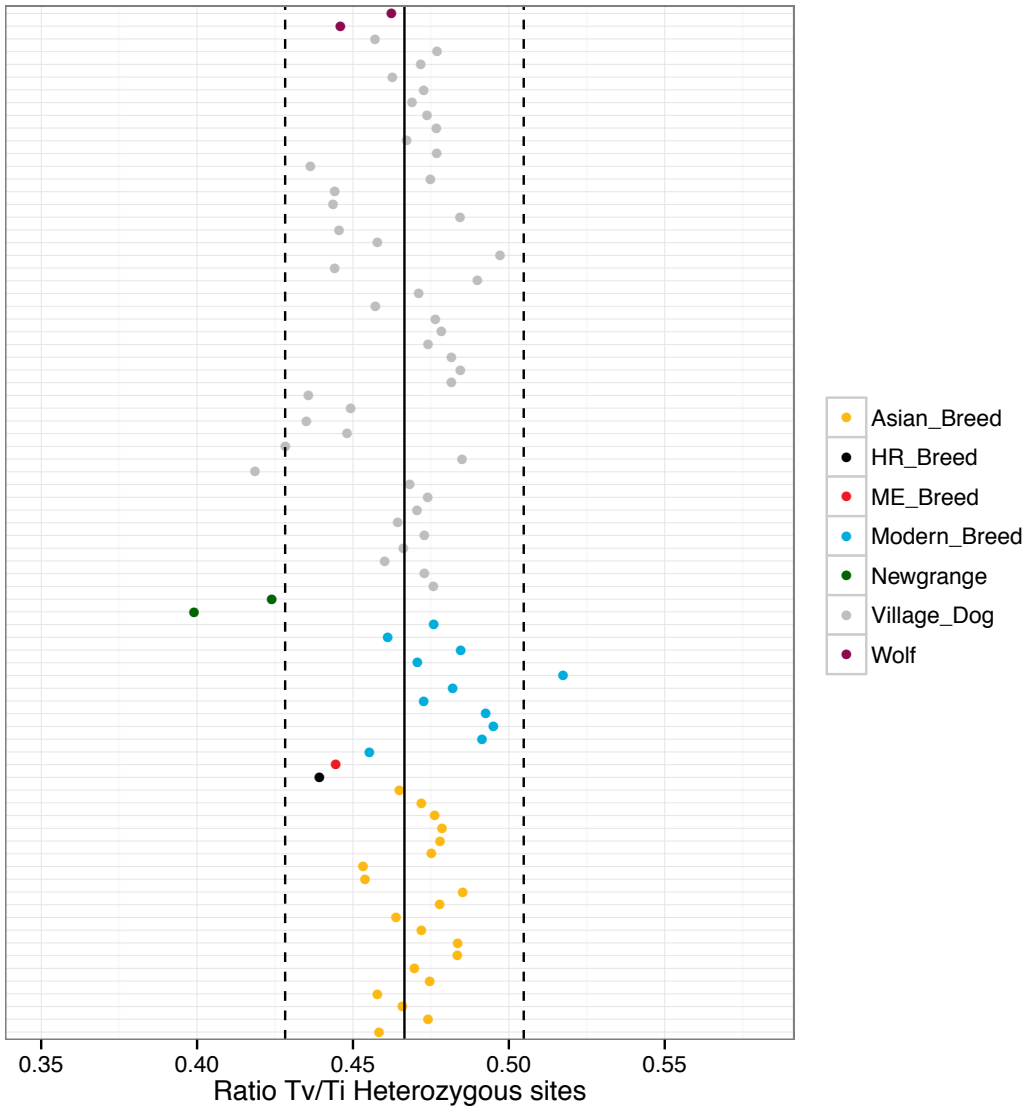


Fig. S26.

Proportion of Tv Ti at heterozygous sites in each genome sample. The two green dots represents re-calibrated (with mapdamage; see above) and non re-calibrated respectively. Solid black line represent mean across all samples, dotted lines represent ±1SD. HR_breed=Istria dog; ME_breed=Xolo; Village dogs = all village dogs; Asian breeds= all Asian breeds (including Kunming dog); Modern breeds= German Shepherd, Tornjack and Labrador.

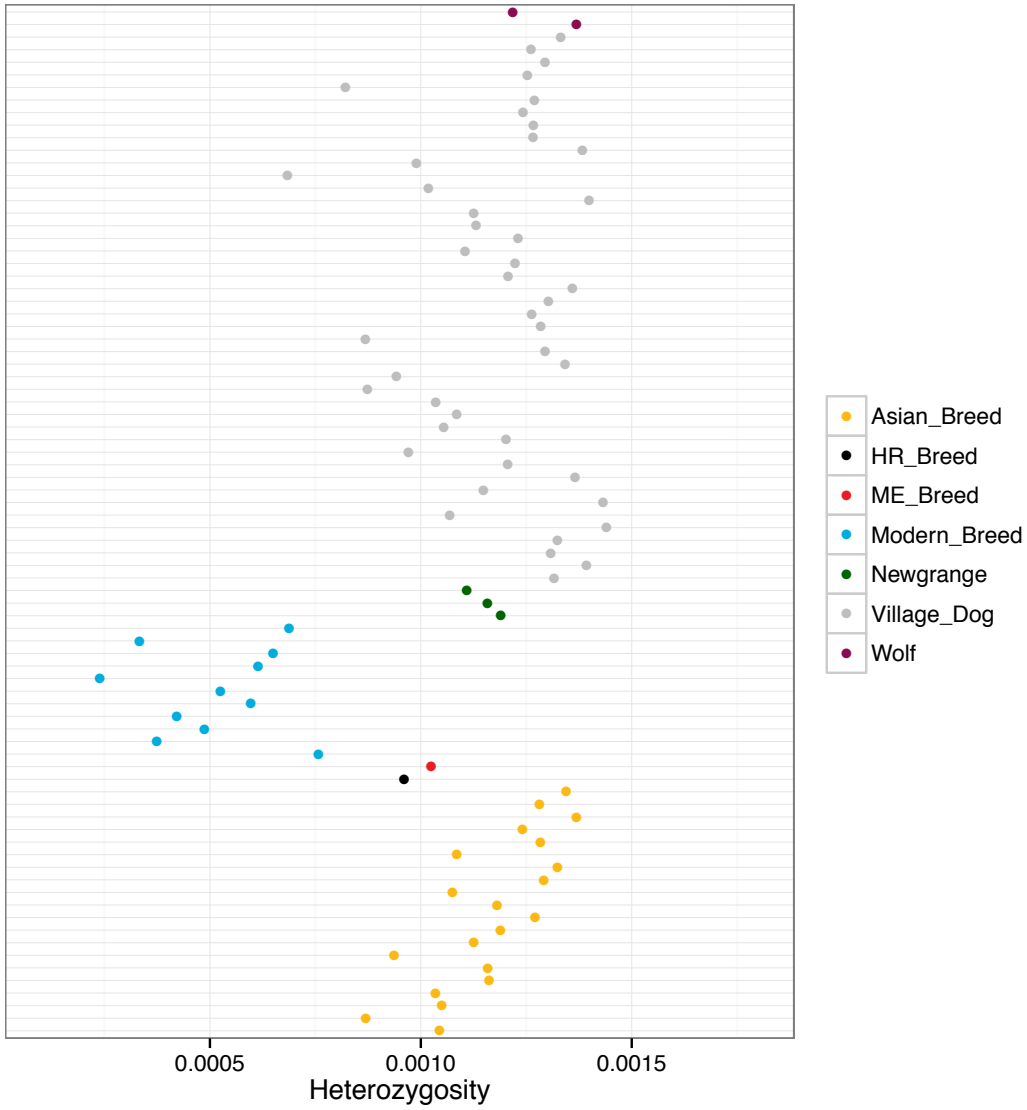


Fig. S27.

Heterozygosity in each genome sample. The two yellow dots represents non re-calibrated, re-calibrated (with mapdamage; see above), corrected (see above) respectively. HR_breed=Istria dog; ME_breed=Xolo; Village dogs = all village dogs; Asian breeds= all Asian breeds (including Kunming dog); Modern breeds= German Shepherd, Tornjack and Labrador.

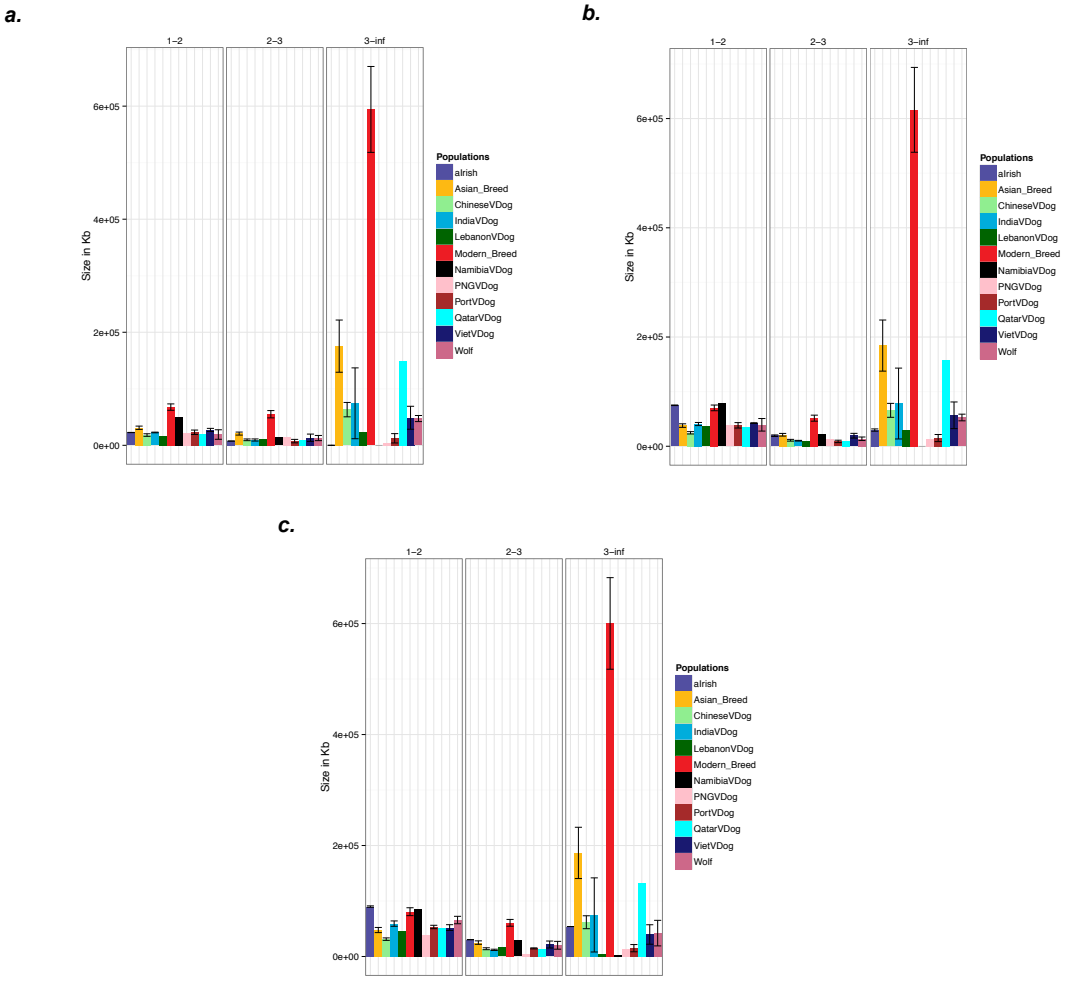


Fig. S28.

Size and occurrence of Runs of homozygosity in genome samples. Error bars represent ±1 SE. ***a.*** One heterozygous site allowed. ***b.*** Two heterozygous sites allowed. ***c.*** Three heterozygous sites allowed.

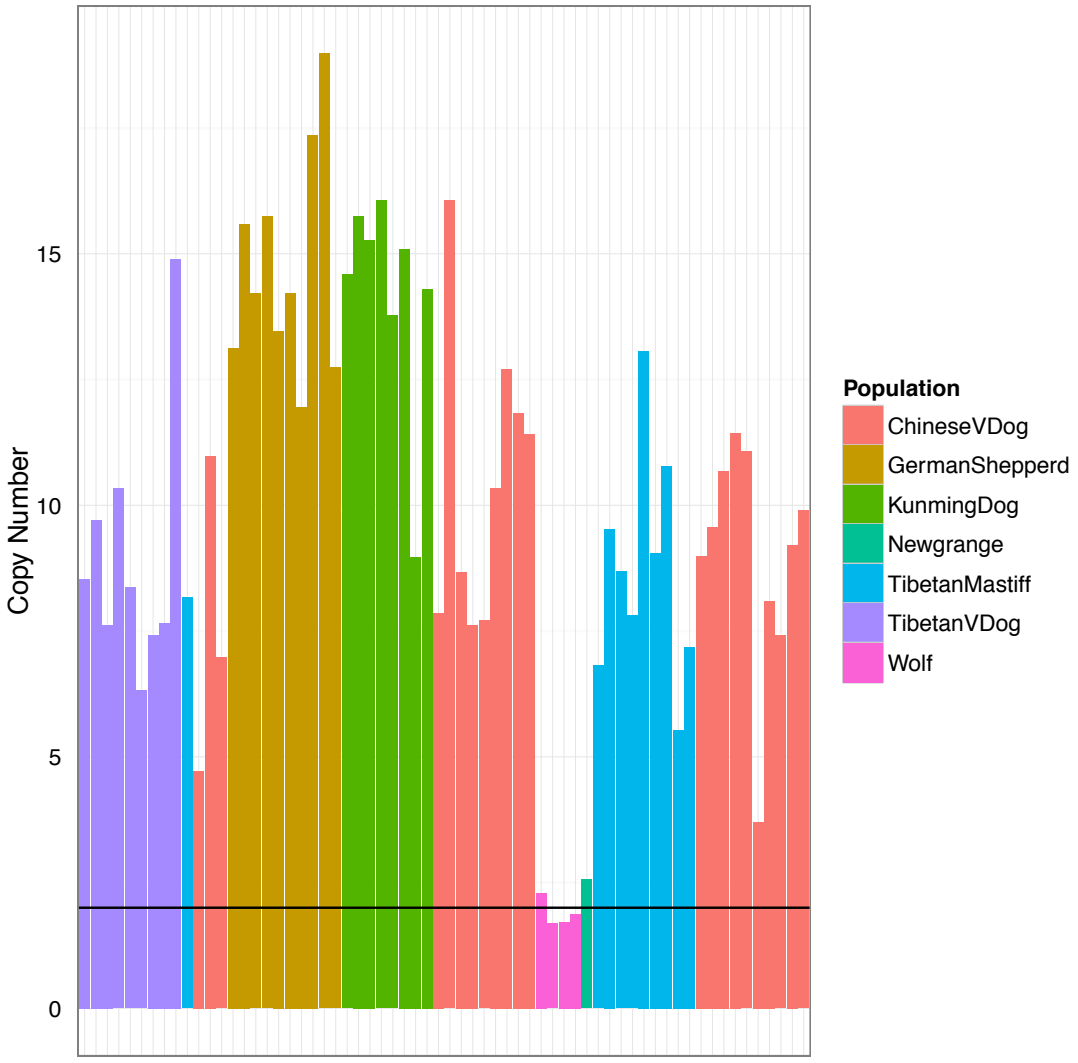


Fig. S29.
Copy number at *AMY2B* gene in various dogs / wolves.

Table S1.
Newgrange dog alignment summary statistics.

Excel file available online.

Table S2.

Summary statistics of the entire set of individuals aligned in this study. #ID: Sample ID; POP: population; DoC: Depth of Coverage; ACCESSION: NCBI accession number Type or paste caption here.

#ID	POP	DoC	ACCESSION
Newgrange	Newgrange	33.6	PRJEB13070
HR85	Modern_Breed	7.3	SRX399726
PG115	PNGVDog	7.0	SRX399745
NA89	NamibiaVDog	7.6	SRX399744
2972	Modern_Breed	10.6	SRX399705
BA19	Modern_Breed	6.9	SRX399708
Dog07	ChineseVDog	6.9	SRX399714
Dog13	ChineseVDog	8.3	SRX399720
QA27	QatarVDog	12.9	SRX399751
IN23	BorneoVDog	7.2	SRX399737
TW04	ChineseVDog	16.9	SRX399754
PT71	PortVDog	16.2	SRX399749
Dog15	ChineseVDog	7.8	SRX399722
LB85	LebanonVDog	10.5	SRX399741
VN37	VietVDog	8.5	SRX399756
HR93	Modern_Breed	11.2	SRX399727
ID168	IndiaVDog	10.0	SRX399733
Dog12	ChineseVDog	8.0	SRX399719
Dog14	ChineseVDog	8.2	SRX399721
VN21	VietVDog	10.9	SRX399755
EG44	EgyptVDog	9.2	SRX399724
Dog04	ChineseVDog	7.4	SRX399712
ID91	IndiaVDog	8.6	SRX399735
1735	Asian_Breed	8.5	SRX399699
NA8	NamibiaVDog	15.2	SRX399743
ID60	IndiaVDog	15.5	SRX399734
IN18	BorneoVDog	5.9	SRX399736
Lcu2	OUTGROUP	12.1	SRX403460,SRX403461
VN76	VietVDog	7.4	SRX399761
PG84	PNGVDog	13.7	SRX399747
PG122	PNGVDog	6.9	SRX399746
Dog10	ChineseVDog	8.2	SRX399717
Dog03	ChineseVDog	8.1	SRX399711

VN42	VietVDog	9.3	SRX399759
Dog08	ChineseVDog	7.9	SRX399715
4669	Modern_Breed	17.0	SRX399707
IN29	BorneoVDog	6.2	SRX399738
Dog02	ChineseVDog	8.4	SRX399710
LB79	LebanonVDog	9.9	SRX399740
Dog09	ChineseVDog	7.5	SRX399716
VN4	VietVDog	10.4	SRX399757
VN59	VietVDog	10.6	SRX399760
ID165	IndiaVDog	11.1	SRX399732
PT61	PortVDog	18.6	SRX399748
LB74	LebanonVDog	8.2	SRX399739
Dog06	ChineseVDog	8.2	SRX399713
NA63	NamibiaVDog	9.1	SRX399742
QA5	QatarVDog	7.5	SRX399753
Dog11	ChineseVDog	7.5	SRX399718
Dog01	ChineseVDog	7.5	SRX399709
EG49	EgyptVDog	8.4	SRX399725
ID125	IndiaVDog	15.2	SRX399728
ID137	IndiaVDog	8.6	SRX399731
GS5	Modern_Breed	15.0	DoGSD
TM10	Asian_Breed	16.4	DoGSD
TM8	Asian_Breed	14.5	DoGSD
KM4	Asian_Breed	16.7	DoGSD
LJ5	ChineseVDog	15.6	DoGSD
DQ9	ChineseVDog	14.8	DoGSD
GS9	Modern_Breed	17.0	DoGSD
KM3	Asian_Breed	16.2	DoGSD
GS8	Modern_Breed	15.4	DoGSD
KM10	Asian_Breed	17.3	DoGSD
YJ10	ChineseVDog	16.0	DoGSD
LJ1	ChineseVDog	17.4	DoGSD
YJ4	ChineseVDog	15.1	DoGSD
TM2	Asian_Breed	14.2	DoGSD
DQ5	ChineseVDog	16.2	DoGSD
LUPWRUS00003	Wolf	12.0	DoGSD
YJ3	ChineseVDog	14.5	DoGSD
LJ6	ChineseVDog	15.6	DoGSD
GS2	Modern_Breed	16.5	DoGSD
GS4	Modern_Breed	15.7	DoGSD
LJ10	ChineseVDog	16.1	DoGSD

ISW	Wolf	6.7	DoGSD
LJ4	ChineseVDog	14.1	DoGSD
DQ6	ChineseVDog	13.7	DoGSD
DQ2	ChineseVDog	16.9	DoGSD
GS10	Modern_Breed	16.1	DoGSD
KM7	Asian_Breed	18.3	DoGSD
DQ8	ChineseVDog	13.4	DoGSD
YJ2	ChineseVDog	15.2	DoGSD
KM1	Asian_Breed	11.8	DoGSD
FAMICHN00001	ChineseVDog	12.8	DoGSD
GS1	Modern_Breed	15.4	DoGSD
TM5	Asian_Breed	15.1	DoGSD
TM3	Asian_Breed	14.5	DoGSD
DQ4	ChineseVDog	15.5	DoGSD
YJ1	ChineseVDog	17.0	DoGSD
DQ1	ChineseVDog	16.3	DoGSD
KM9	Asian_Breed	14.8	DoGSD
LUPWCHN00001	Wolf	9.2	DoGSD
YJ8	ChineseVDog	15.7	DoGSD
KM5	Asian_Breed	16.5	DoGSD
TM9	Asian_Breed	16.4	DoGSD
TM6	Asian_Breed	15.6	DoGSD
CRW	Wolf	8.7	DoGSD
LJ8	ChineseVDog	16.3	DoGSD
YJ7	ChineseVDog	15.3	DoGSD
TM1	Asian_Breed	16.2	DoGSD
GS7	Modern_Breed	16.9	DoGSD
KM6	Asian_Breed	16.3	DoGSD
GS3	Modern_Breed	15.9	DoGSD
KM8	Asian_Breed	14.5	DoGSD
TM7	Asian_Breed	16.3	DoGSD
FAMICHN00002	ChineseVDog	10.8	DoGSD
FAMBGSD00001	Modern_Breed	9.4	DoGSD
FAMBTIM00001	Asian_Breed	11.0	DoGSD
DQ3	ChineseVDog	14.9	DoGSD
Dingo	Dingo	6.7	DoGSD
LJ7	ChineseVDog	13.4	DoGSD
Basenji	Basenji	6.3	DoGSD
YJ5	ChineseVDog	14.3	DoGSD
YJ9	ChineseVDog	15.3	DoGSD
GLJ	ChineseVDog	5.3	DoGSD

TM4	Asian_Breed	13.7	DoGSD
GS6	Modern_Breed	16.3	DoGSD
LJ2	ChineseVDog	17.1	DoGSD
FAMICHN00003	ChineseVDog	9.4	DoGSD
LUPWRUS00001	Wolf	11.3	DoGSD
DQ7	ChineseVDog	17.1	DoGSD
LUPWRUS00002	Wolf	9.1	DoGSD
LJ9	ChineseVDog	16.8	DoGSD
YJ6	ChineseVDog	15.1	DoGSD

Table S3.

Total number of SNPs called in 80 individuals. NHOM= Number of homozygous non-reference sites; NHET= Number of heterozygous; NHOM_TI= Number of homozygous transitions; NHOM_TV=Number of homozygous transvertions; NHET_TI=Number of heterozygous transitions; NHET_TV=Number of heterozygous transversions.

#ID	#NSNP	#NHOM	#NHET	#NHOM_TI	#NHOM_TV	#NHET_TI	#NHOM_TV
2972	2189944	1081679	1108265	352658	729021	346530	761735
4669	3522663	1305444	2217219	426594	878850	675907	1541312
DQ1	4533114	1863388	2669726	631532	1231856	862018	1807708
DQ2	4633528	1851930	2781598	628178	1223752	898828	1882770
DQ3	4411051	1642069	2768982	553800	1088269	881315	1887667
DQ4	4370555	1809161	2561394	608576	1200585	816657	1744737
DQ5	4555141	1933761	2621380	653260	1280501	845527	1775853
DQ6	4194086	1765572	2428514	592695	1172877	774663	1653851
DQ7	4525441	1838639	2686802	621051	1217588	862264	1824538
DQ8	4022551	1714238	2308313	575556	1138682	742933	1565380
DQ9	4315768	1639578	2676190	556058	1083520	862098	1814092
FAMBTIM00001	3191169	1347037	1844132	441041	905996	577209	1266923
FAMICHN00001	3183352	1201669	1981683	392979	808690	589468	1392215
FAMICHN00002	2436353	1285058	1151295	423507	861551	375829	775466
GS10	3607111	1958315	1648796	655167	1303148	531639	1117157
GS1	3550946	1951018	1599928	655322	1295696	514561	1085367
GS2	3692390	1969736	1722654	662015	1307721	554651	1168003
GS3	3602346	1961801	1640545	659333	1302468	527008	1113537
GS4	3497323	2078990	1418333	696211	1382779	456893	961440
GS5	3305178	2045239	1259939	683787	1361452	413506	846433
GS6	3634625	2029111	1605514	681438	1347673	514195	1091319
GS7	3711193	1976218	1734975	664912	1311306	558852	1176123
GS9	3692298	1943131	1749167	655458	1287673	565813	1183354
HR93	2407807	1040923	1366884	336187	704736	418125	948759
ID125	3588492	1414960	2173532	453146	961814	654022	1519510
ID165	3036458	1201509	1834949	390737	810772	565386	1269563
ID168	2544990	1045322	1499668	331827	713495	451393	1048275
ID60	3692037	1735220	1956817	551959	1183261	597550	1359267
KM10	3884377	1266514	2617863	444823	821691	844701	1773162
KM1	3129107	1737313	1391794	578602	1158711	444323	947471

KM3	3740778	1939911	1800867	652860	1287051	577473	1223394
KM4	3935683	1825211	2110472	619234	1205977	672738	1437734
KM5	3993908	1754865	2239043	594121	1160744	717896	1521147
KM6	3832550	1357968	2474582	472139	885829	795427	1679155
KM7	4135378	1247623	2887755	439466	808157	930733	1957022
KM8	3736204	1625020	2111184	549180	1075840	679629	1431555
KM9	3763916	1474691	2289225	503986	970705	732211	1557014
LB85	2162779	875695	1287084	283677	592018	395864	891220
LJ10	4022971	1440089	2582882	499636	940453	857054	1725828
LJ1	4091316	1693865	2397451	582519	1111346	799378	1598073
LJ2	4275249	1729592	2545657	591049	1138543	835116	1710541
LJ4	4285869	1782255	2503614	601897	1180358	802462	1701152
LJ5	4021533	1421971	2599562	490930	931041	849582	1749980
LJ6	4403763	1743680	2660083	590254	1153426	853982	1806101
LJ7	4158749	1457007	2701742	494852	962155	853658	1848084
LJ8	4346452	1574220	2772232	538289	1035931	892506	1879726
LJ9	3876473	1572316	2304157	541447	1030869	765057	1539100
LUPWRUS00001	3806917	1632533	2174384	525819	1106714	670011	1504373
LUPWRUS00003	3837170	1695801	2141369	554750	1141051	683666	1457703
NA8	3494990	1362045	2132945	444415	917630	654898	1478047
Lcu2	18424833	16585646	1839187	5005347	11580299	568121	1271066
Newgrange	4068176	1510901	2557275	500487	1010414	708724	1848551
Newgrange_rescaled	3945507	1501977	2443530	500859	1001118	709639	1733891
PG84	2903678	1080985	1822693	372244	708741	605271	1217422
PT61	3776863	1404247	2372616	469759	934488	741232	1631384
PT71	3485385	1304715	2180670	429742	874973	670128	1510542
QA27	3227516	1299119	1928397	445552	853567	632788	1295609
TM10	4561268	1832272	2728996	618195	1214077	871656	1857340
TM1	4313326	1902304	2411022	641177	1261127	780961	1630061
TM2	4352845	1664378	2688467	556716	1107662	848443	1840024
TM3	4329138	1718408	2610730	573898	1144510	820424	1790306
TM4	3951366	1666601	2284765	561086	1105515	737476	1547289
TM5	4182569	1625822	2556747	553434	1072388	834112	1722635
TM6	4171491	1693082	2478409	574255	1118827	807625	1670784
TM7	4623251	1746029	2877222	594127	1151902	934020	1943202
TM8	4154196	1719542	2434654	580112	1139430	784995	1649659

TM9	4540862	1870558	2670304	631736	1238822	852237	1818067
TW04	3999168	1388648	2610520	451879	936769	795045	1815475
VN21	2707473	1241577	1465896	397371	844206	454387	1011509
VN4	2907183	1391119	1516064	453364	937755	482887	1033177
VN59	2710746	1239722	1471024	390541	849181	446264	1024760
YJ10	4556142	1848455	2707687	623522	1224933	867598	1840089
YJ1	4553050	1769872	2783178	601176	1168696	901330	1881848
YJ2	4131511	1707322	2424189	576545	1130777	792766	1631423
YJ3	4255270	1780514	2474756	600639	1179875	802228	1672528
YJ4	4151555	1858117	2293438	622937	1235180	741485	1551953
YJ5	4102489	1747558	2354931	587054	1160504	757243	1597688
YJ6	4437692	1721627	2716065	580525	1141102	868889	1847176
YJ7	4275072	1781874	2493198	599429	1182445	807546	1685652
YJ8	4377493	1771227	2606266	598391	1172836	842097	1764169
YJ9	4438877	1716654	2722223	578128	1138526	867395	1854828

Table S4.
D-statistics results from whole genome.

Excel file available online.

Table S5.
Summary of the mtDNA data set.

Excel file available online.

Table S6.

Copy number and coordinates of the AMY2B gene in the boxer reference genome (canfam3.1).

Chrom	Score	Match	Strand	Start	End	Length
Un_AAEX03024353	224	100.00%	+	2812	3035	224
Un_AAEX03022739	224	100.00%	+	9681	9904	224
Un_AAEX03020568	224	100.00%	+	19963	20186	224
Un_AAEX03020568	224	100.00%	+	34785	35008	224
Un_AAEX03020568	224	98.70%	+	5109	5332	224

Table S7.

Archeological information plotted in Fig. 3a. Samples with a * represent “proto-dogs” specimens that were excluded from the analysis (see above).

Excel file available online.

Supplementary References:

23. C. Gamba *et al.*, *Nat. Commun.* **5**, 5257 (2014).
24. L. H. van Wijngaarden-Bakker, *Proc. R. Irish Acad. Sect. C Archaeol. Celt. Stud. Hist. Linguist. Lit.* **86C**, 17–111 (1986).
25. M. J. O'Kelly, R. M. Cleary, D. Lehane, *Newgrange, Co. Meath, Ireland: The Late Neolithic/Beaker Period Settlement* (B.A.R., 1983).
26. F. McCormick, in *The Holocene History of the European Vertebrate Fauna: Modem Aspects of Research*, N. Benecke, Ed. (1998).
27. P. Reimer, *Radiocarbon*. **55**, 1869–1887 (2013).
28. S. Pääbo *et al.*, *Annu. Rev. Genet.* **38**, 645–679 (2004).
29. L. Orlando *et al.*, *Genome Res.* **21**, 1705–19 (2011).
30. J. T. Vilstrup *et al.*, *PLoS One*. **8**, e55950 (2013).
31. M. Meyer, M. Kircher, *Cold Spring Harb. Protoc.* **2010**, pdb.prot5448 (2010).
32. M. Martin, *EMBnet.journal*. **17**, 10 (2011).
33. H. Li, R. Durbin, *Bioinformatics*. **25**, 1754–60 (2009).
34. K. Lindblad-Toh *et al.*, *Nature*. **438**, 803–19 (2005).
35. M. Schubert *et al.*, *Nat. Protoc.* **9**, 1056–82 (2014).
36. H. Li *et al.*, *Bioinformatics.* **25**, 2078-9 (2009).
37. A. McKenna *et al.*, *Genome Res.* **20**, 1297–1303 (2010).
38. K. Okonechnikov, A. Conesa, F. García-Alcalde, *Bioinformatics*. **32**, 292-4 (2015).
39. H. Jónsson, A. Ginolhac, M. Schubert, P. L. F. Johnson, L. Orlando, *Bioinformatics*. **29**, 1682–4 (2013).
40. B. Bai *et al.*, *Nucleic Acids Res.* **43**, 777-83 (2014).
41. A. Auton *et al.*, *Science*. **336**, 193-8 (2012).
42. A. R. Quinlan, I. M. Hall, *Bioinformatics*. **26**, 841–2 (2010).
43. E. Han, J. S. Sinsheimer, J. Novembre, *Mol. Biol. Evol.* **31**, 723–35 (2014).
44. A. Vaysse *et al.*, *PLoS Genet.* **7**, e1002316 (2011).
45. M. L. Speir *et al.*, *Nucleic Acids Res.*, **44** 717–25 (2016).
46. S. Purcell *et al.*, *Am. J. Hum. Genet.* **81**, 559–75 (2007).
47. E. Paradis, J. Claude, K. Strimmer, *Bioinformatics*. **20**, 289–290 (2004).
48. N. Patterson, A. L. Price, D. Reich, *PLoS Genet.* **2**, e190 (2006).
49. J. K. Pickrell, J. K. Pritchard, *PLoS Genet.* **8**, e1002967 (2012).

50. E. Y. Durand, N. Patterson, D. Reich, M. Slatkin, *Mol. Biol. Evol.* **28**, 2239–2252 (2011).

51. N. Patterson *et al.*, *Genetics*. **192**, 1065–93 (2012).

52. S. R. Browning, B. L. Browning, *Am. J. Hum. Genet.* **81**, 1084–97 (2007).

53. Z. Fan *et al.*, *Genome Res.* **26**, 163-73 (2016).

54. H. Angleby, P. Savolainen, *Forensic Sci. Int.* **154**, 99–110 (2005).

55. A. Ardalan *et al.*, *Ecol. Evol.* **1**, 373–85 (2011).

56. A. R. Boyko *et al.*, *Proc. Natl. Acad. Sci. U. S. A.* **106**, 13903–8 (2009).

57. S. K. Brown *et al.*, *PLoS One*. **6**, e28496 (2011).

58. S. Castroviejo-Fisher, P. Skoglund, R. Valadez, C. Vilà, J. A. Leonard, *BMC Evol. Biol.* **11**, 73 (2011).

59. V. Muñoz-Fuentes, C. T. Darimont, P. C. Paquet, J. A. Leonard, *Conserv. Genet.* **11**, 547–556 (2009).

60. N. Okumura, N. Ishiguro, M. Nakano, A. Matsui, M. Sahara, *Anim. Genet.* **27**, 397–405 (2009).

61. M. C. R. Oskarsson *et al.*, *Proc. Biol. Sci.* **279**, 967–74 (2012).

62. J.-F. Pang *et al.*, *Mol. Biol. Evol.* **26**, 2849–64 (2009).

63. P. Savolainen, T. Leitner, A. N. Wilton, E. Matisoo-Smith, J. Lundeberg, *Proc. Natl. Acad. Sci. U. S. A.* **101**, 12387–90 (2004).

64. K. Tsuda, Y. Kikkawa, H. Yonekawa, Y. Tanabe, *Genes Genet. Syst.* **72**, 229–38 (1997).

65. J. A. Leonard *et al.*, *Science*. **298**, 1613–6 (2002).

66. M. F. Deguilloux, J. Moquel, M. H. Pemonge, G. Colombeau, *J. Archaeol. Sci.* **36**, 513–519 (2009).

67. F. Verginelli *et al.*, *Mol. Biol. Evol.* **22**, 2541–51 (2005).

68. M. Pionnier-Capitan, thesis, ENS Lyon (2010).

69. O. Lebrasseur, thesis, Durham University (2014).

70. N. Rohland, M. Hofreiter, *Nat. Protoc.* **2**, 1756–62 (2007).

71. A. Cooper, *Science (80-.).* **289**, 1139b–1139 (2000).

72. M. T. P. Gilbert, H.-J. Bandelt, M. Hofreiter, I. Barnes, *Trends Ecol. Evol.* **20**, 541–4 (2005).

73. K. S. Kim, S. E. Lee, H. W. Jeong, J. H. Ha, *Mol. Phylogenet. Evol.* **10**, 210–20 (1998).

74. M.-S. Peng *et al.*, *Mol. Ecol. Resour.* **15**, 1238–42 (2015).

75. R. C. Edgar, *Nucleic Acids Res.* **32**, 1792–7 (2004).

76. J. Leigh, PopART (Population Analysis with Reticulate Trees) (2015).

77. R. R. Hudson, *Bioinformatics*. **18**, 337–338 (2002).

78. G. K. Chen, P. Marjoram, J. D. Wall, *Genome Res.* **19**, 136–42 (2009).

79. D. H. Alexander, J. Novembre, K. Lange, *Genome Res.* **19**, 1655–64 (2009).

80. M. Boudadi-Maligne, J.-B. Mallye, M. Langlais, C. Barshay-Szmidt, *PALEO*, 39–54 (2012).

81. R. J. Losey *et al.*, *PLoS One*. **8**, e63740 (2013).

82. N. Benecke, *J. Archaeol. Sci.* **14**, 31–49 (1987).

83. M. Germonpré *et al.*, *J. Archaeol. Sci.* **36**, 473–490 (2009).

84. R. Flad, *J. East Asian Archaeol.* **3**, 23–51 (2001).

85. L. Barton *et al.*, *Proc. Natl. Acad. Sci. U. S. A.* **106**, 5523–8 (2009).

86. A. Lasota-Moskalewska, K. Szymczak, M. Khudzhanazarov, *Archaeol. Balt.* **11** (2009).

87. A. Razzokov. *Sarazm. Dushanbe: Institute of History, Archaeology, and Ethnography. Dushanbe, Tajikistan: Academy of Sciences of Tajikistan A. Donish History, Archaeology and Ethnographic Institute.* (2008).

88. H. Bocherens, M. Mashkour, D. Billiou, *Environ. Archaeol.* **33**, 253-64 (2013).

89. A. K. Kasparov, *Paléorient*. **22**, 161–167 (1996).

90. M. Frachetti, N. Benecke, *Antiquity*. **83**, 1023–1037 (2015).

91. C. Chang, N. Benecke, F. Grigoriev, A. Rosen, P. Tourtellotte, *Antiquity.* **77**, 298-312 (2003)

92. V. Bakhshaliev, in *The Archaeology of Nakhichevan. Ten Years of New Discoveries* (Ege Yayinlari, Istanbul, 1997).

93. R. Berthon *et al.*, *Environ. Archaeol.* **18**, 191-200 (2013).

94. S. J. Pawankar and P. K. Thomas, *Bulletin of the Deccan College Research Institute* **56**, 363-370 (1996).

95. B. & R. Allchin, F. R. Allchin, in *The Birth of Indian Civilization: India and Pakistan Before 500 B.C.* (1968).

96. R. V. Joshi, *Archaeol. Surv. India, New Delhi, India*, 12–16 (1961).

97. P. K. Thomas, in *The Walking larder: patterns of domestication, pastoralism, and predation* (Routledge, 2014).

98. B. P. Sahu, *From hunters to breeders: faunal background of early India* (Anamika Prakashan, 1988).

99. N. Bhola, G. V. S. Rao, *Animal remains from Lothal excavations* (Zoological Survey of India, 1962).

100. R. B. . Sewell, B. S. Guha, in *Mohenjo-Daro and the Indus* (1931).

101. G. Stein, in *Early Animal Domestication and Its Cultural Context*, P. J. Crabtree, D. V. Campana, K. Ryan, Eds. (UPenn Museum of Archaeology, 1989).

102. Telegin D.Ya., *redno-Stogivska kultura epokhi midi.* (1973).

103. K. N. Wilkinson *et al.*, *J. F. Archaeol.* **37**, 20-33 (2013).

104. R. D. Barnett, W. Watson, *Iraq*. **14**, 132 (1952).

105. M. D. Frachetti, N. Benecke, A. N. Mar'yashev, P. N. Doumani, *World Archaeol.* **42**, 622–646 (2010).

106. P. M. Dolukhanov, in *Hunters in transition: Mesolithic societies of temperate Eurasia and their transition to farming* (1986).

107. G. Matyushin, *The mesolithic and neolithic in the southern Urals and central Asia.* (1986).

108. N. D. Ovodov *et al.*, *PLoS One*. **6**, e22821 (2011).

109. M. Germonpré, M. Lázničková-Galetová, R. J. Losey, J. Räikkönen, M. V. Sablin, *Quat. Int.* **359-360**, 261–279 (2015).

110. M.-A. Garcia, *Bull. la Société préhistorique française*. **102**, 103–108 (2005).

111. M. V. Sablin, G. A. Khlopachev, *Curr. Anthropol.* **43**, 795–799 (2002).

112. A. Perri, *J. Archaeol. Sci.* **68**, 1–4 (2016).

113. S. J. Crockford, Y. V. Kuzmin, *J. Archaeol. Sci.* **39**, 2797–2801 (2012).

114. M. Boudadi-Maligne, G. Escarguel, *J. Archaeol. Sci.* **45**, 80–89 (2014).

115. A. G. Drake, M. Coquerelle, G. Colombeau, *Sci. Rep.* **5**, 8299 (2015).

116. D. P. Howrigan, M. A. Simonson, M. C. Keller, *BMC Genomics*. **12**, 460 (2011).

117. E. Axelsson *et al.*, *Nature*. **495**, 360–4 (2013).

118. E. Cadieu *et al.*, *Science*. **326**, 150–3 (2009).

119. H. G. Parker *et al.*, *Science*. **325**, 995–8 (2009).